2018

QUANTITATIVE AND MOLECULAR ANALYSIS OF HABITUATION AT THE MAIZE r1 LOCUS

Robert C. Lindsay
Integrated Life Sciences PhD Program

QUANTITATIVE AND MOLECULAR ANALYSIS OF HABITUATION AT THE MAIZE *r1* LOCUS

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

By
ROBERT CLARK LINDSAY
BACHELOR OF ART, HOOD COLLEGE 2008
MASTER OF SCIENCE, VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY 2010

Director: WILLIAM B. EGGLESTON, JR., PHD
ASSOCIATE PROFESSOR, DEPARTMENT OF BIOLOGY

Virginia Commonwealth University
Richmond, Virginia
November 2018

Acknowledgments

I would like to acknowledge several people, but especially the help and guidance of Dr. Eggleston. I would like to thank my wife, Tina, for her help and understanding in the project, and I would like to thank the many undergraduate volunteers who were kind enough to help me: Erica Street, Damien Johnson, Dave Sani, Louis Akinola, Mason Gerena, Monica Evbadoloyi, Michelle Heinz, Areeg Ibdah, Spencer Shin, Emily Pennington, Mason Gerena, and David Buenrostro. I would like to thank Joy Sabl for her technical assistance. I would also like to thank my committee members, Dr. Frenando Tenjo, Dr. Shirley Taylor, Dr. Jennifer Stewart, and Dr. Rita Shiang. I would also like to thank the Integrated Life Sciences Program for summer funding, and for several years of funding as a teaching assistant.

And finally,

For Malcolm

Table of Contents

Clarification of Contributions:

Science is a collaborative effort, and during my training as a graduate student, I received help from many people. I am forever grateful for their assistance in this project, and could not have done it without them.

Chapter 2:

Kernel sample cleaning and light reflectometry sampling was conducted by the following undergraduates: Michelle Heinz, Areeg Ibdah, Spencer Shin, Emily Pennington, Mason Gerena, Louis Akinola, and David Buenrostro. Special thanks to Louis Akinola, who cross-checked the all of the data entry of the sample reads.

Chapter 3:

Damien Johnson, Erica Street, Dave Sani, and Monica Evbadoloyi assisted in DNA extraction and sample preparation for sequencing. Special thanks to Monica Evbadoloyi and Damien Johnson, who helped develop the crown kernel DNA extraction protocol.

Chapter 4:

Joy Sabl assisted greatly in design and theory of the degenerate primers for sodium bisulfite PCR.

List of Tables:

List of Figures:

List of Abbreviations and Definitions

Crown - the top third of the maize kernel

Ear - one ear of corn from a plant

Family - group of plants that originate from the same parent plant andear

IR - Inhibitor of R

Hsp90 - Heat Shock Protein 90 kilodaltons

Line - allele, such as *R-sc:86*

Mop1 - Mediator of Paramutation gene

Nc gene - Near Colorless gene

r1 gene - Maize *red color1* gene

rmr6- - required to maintain suppression 6 gene

R-st - R Stippled line

Sample - all of the shelled kernels from a single ear

Subline - a derivative of a line, *R-sc:86-17pale* Lightest

T Cells - a lymphocyte active in the human immune system

TE - Transposable Element

Abstract

MOLECULAR ANALYSIS OF LOSS OF CANALIZATION AND HABITUATION AT THE MAIZE *r1* LOCUS

By Robert Clark Lindsay, B.A., M.hS.

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Major Advisor: William B. Eggleston, Jr.
Associate Professor, Department of Biology

Epigenetics is the study of heritable changes in phenotypes that are not the result of changes in DNA sequence. Examples of epigenetic affecters include methylation changes, chromatin modifications, transcription factors, and RNA-based changes. The molecular mechanisms behind epigenetic changes are not fully understood. Canalization is the buffering of gene expression against environmental changes over time, while habituation is semi-stable expression change over time due to selection. This work characterized the molecular changes associated with the kernel color changes of the *R-sc:86-17pale* allele at the maize *red color1* (*r1*) locus to determine if the changes are epigenetic in nature. The research; 1) quantified the color differences between the progenitor and habituated sublines; 2) Determined that there are not sequence differences between the progenitor and habituated sublines at the 3` end of the *Sc*||nc1 gene that could account for changes in seed color; 3) and examined the cytosine methylation patterns at the 3` end of the *Sc*||*nc1* gene of the habituated sublines and the progenitor to determine whether there are methylation differences that correspond with the kernel color changes. Quantification of the kernel colors of the *R-sc:86*-17pale selection sublines

showed that there was a statistically significant difference in kernel color. The identical sequence of the *R-sc:86* line and the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines at the 3` end of the *Sc||nc1* gene is evidence that the kernel color change is not driven by differences in sequence within the *r1* gene. The methylation data suggests that some methylation differences in the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines are present, and suggests that the molecular basis of the kernel color is epigenetic in nature.

Chapter 1

Epigenetics, the *r1* locus, and the indentification of the *R-sc:86-17pale* Selection sublines

Epigenetics:

In 1942, Waddington described epigenetics, which is the mechanism of heritable expression and development patterns that are not caused by genetic changes. Recent work has shown that epigenetic phenomena resulting in changes in gene expression include: cytosine methlyation, modification of chromatin and histones, transcription factors, and RNA modification/regulation (Freeling and Lisch 2010, Erhard and Hollick 2011, Wang 2011, Ponomarev *et al.* 2012*,* Dunn *et al.* 2015*,* Bustos *et al.* 2017, Cifuentes-Zuniga *et al.* 2017, Kishi *et al.* 2017, Salehi 2017). Epigenetic changes can directly affect transcription, translation, RNA processing and protein modification, which in turn alter phenotypes, and have been found to occur in both single and multigenic traits in single and multicellular organisms (Wang *et al.* 2009, Eichten *et al.* 2011).

Epigenetics can involve heritable changes in control of gene expression, often in response to environmental pressures, as can be seen in the downstream effects of influenza virus infections and in changes in human expression of traits in response to famine (Bedford 2012, Bygren *et al.* 2014, Tang *et al.* 2011, Dunn 2015). These epigenetic changes can take the form of affecting nutrient metabolism in the grandchildren of famine survivors, as well as changes in T cell expression patterns in response to influenza (Bygren *et al.* 2014, Dunn 2015). This effect of environmental pressures affecting heritable fitness revives the Lamarckian idea that evolution is not solely based on Mendelian genetics and the occurrence and movement of

alleles/epialleles through a population, but is also dependent on epigenetic changes in gene expression (Fanti *et al.* 2017). This leads to a need to understand epigenetic regulation of gene expression through such phenomena as canalization and habituation.

Canalization:

Canalization, or the channeling of expression patterns into a few phenotypic patterns in spite of allelic variation or environmental pressure, was first described by Waddington, who hypothesized a need to maintain normal expression patterns until abnormal environmental factors lead to the expression of cryptic alleles (Figure 1) (Waddington 1956, Waddington 1942). These normal expression patterns were described as 'canals' that directed the expression to normal, or 'regular' phenotypes, masking expression of abnormal alleles and traits (cryptic variation). Canalization of a phenotype stabilizes expression of the trait, reducing phenotypic variation (narrow expressivity), and potentially optimizing expression of the trait for the current environmental selection pressures (Figure 2) (Hornstein & Shomron 2006, Salathia & Queitsch 2007). Chanderbali *et al.* (2010) showed that plant species can have different levels of canalization by comparing carpel and tepal lengths to stamen length in *Arabidopsis thaliana* and *Persea americana*. *P. americana* displays significantly less variation then *Arabidopsis thaliana* in the tepal to stamen length ratio, indicated by higher r values and less variation in Figure 2. This lower variation indicates a significantly narrower range of the expression of traits affecting tepal and stamen length than displayed by *A. thaliana*, which indicates stronger canalization of these traits in *P. americana* (Figure 2). Meins and Lutz (1979) proposed that the standardization of gene expression patterns provides

phenotypic stability while retaining epialleles that may allow for different expression patterns in different environments. This standardization of gene expression can be regulated by *Hsp90 (Heat Shock Protein 90*), which can mask expression of epialleles that allow for different expression patterns (cryptic alleles) (Figure 3) (Sato & Siomi 2010). The ability to retain the unexpressed epialleles is postulated to be an important aspect of an organism's ability to react to environmental changes in a way that allows for survival, using alleles that may not normally be selected for, such as in some instances of bacterial drug resistance via plasmids (Martinez & Baquero 2000, Marquez *et al.* 2011, Day 2016). The mechanism(s) underlying canalization are poorly understood (Fanti *et al.* 2017, Sieriebriennikov *et al.* 2017, Carey *et al.* 2015, Le Roy *et al.* 2017, McNamara *et al.* 2016). Both miRNA and reduction of *Hsp90* expression have been found to affect canalized phenotypes, affecting many genes and pathways at the same time (Hornstein & Shomron 2006, Sato & Siomi 2010, Sieriebriennikov 2017). Perturbations of canalization, such as those involving the pathway including *Hsp90,* appear to affect global canalization throughout the organism, rather than working on an individual gene basis (Bedford 2012). As Bogdan *et al.* (2017) proposed, *Hsp90* may help maintain canalization and reduce expression of cryptic phenotypes while active. While canalization may be affected by miRNA, and siRNA in *Arabidopsis*, current research suggests that the mechanism of canalization is not interfering RNA-based, but may be connected with methylation states (Fanti *et al.* 2017, Le Roy 2017, McNamara 2016, Milton 2006).

Loss of canalization results in greater phenotypic variation (wide expressivity) reviewed by Sieriebriennikov (2017) (Rutherford & Lindquist, 1998). Sato and Siomi

(2010) described the buffering role of canalization as channeling expression into only a few expressed phenotypes (black peaks) (Figure 3). This buffering, when reduced or removed, allows the number of expressed phenotypes to increase, as the reduced buffering allows already present cryptic alleles (grey peaks) to be expressed despite their low penetrance. This change in expression patterns allow the total number of different expressed phenotypes to increase, as both cryptic and normally expressed alleles are expressed, without a requirement for a new mutation to drive expression of each phenotypic change. An example of a loss or reduction in canalization is the phenotypic variations that arise following inhibition of the Hsp90 in *Drosophila* and *Arabidopsis thaliana* (Figure 4 & 5) (Queitsch 2002). The Hsp90 protein chaperons protein folding, protein degradation, and stabilizes proteins against heat stress (Salathia & Queitsch 2007). When Hsp90 protein function in *Drospohilia* is disrupted, canalization is reduced, resulting in simultaneous increased phenotypic variation in expression in multiple traits, including eye and body color (Figure 5) (Salathia & Queitsch 2007, Sato & Siomi 2010, Rutherford and Lindquist 1998, Queitsch *et al.* 2002). These phenotypic changes are not limited to organisms with *Hsp90* loss of function mutations, but do occur at a much higher rate than they would normally in wild type animals (Figure 5). Inactivation of *Hsp90* cannot currently be used for directed modification of specific traits due to the global effects of *Hsp90* loss on multiple traits or multiple genes in a pathway, such as can be seen in the anthrocyanin pathway, because there is a high probability of undesirable traits being affected at the same time (pleiotropy), or of the desired trait being masked by off target phenotypes, complicating the selection of the desired phenotype (Fanti *et al.* 2017, Salathia & Queitsch 2007).

Habituation:

Habituation results from heritable changes in the epigenetic state of cells, which occur gradually, are reversible, and do not require genetic changes underlying the visible phenotypic change (Meins & Lutz 1979, Meins 1989). The habituation documented by Meins and Lutz (1979) allowed progeny cells to survive levels of cytokinins that were lethal to un-habituated tobacco progenitor cells after 3 generations of selection, and in which the cytokinin tolerance after three more cell generations could be reversed. The rate of change in cellular cytokinin requirements reported by Meins and Lutz (1979) is 100 - 1000 times faster than can be accounted for by normal mutation rates of 2.9×10^{-8} mutations/base per replication, which coupled with the reversibility of the cytokinin requirements, indicates that habituation likely is not dependant on genetic mutations for changes in phenotype (Meins & Lutz 1979, Clark *et al.* 2005, Largo-Gosens *et al.* 2016, de Castro *et al.* 2017). Habituation can be found in humans with drug habituation though tolerance to increasing dosages of drugs, and in some aspects reminiscent of Lamarkian evolution, such as can be seen in generational increases in heavy metals tolerance in plants (Carey *et al.* 2015, de Castro *et al.* 2017, Macnair 1987).

The maize *r1* locus:

The maize *r1* (red color 1) locus is one of several loci that regulate anthrocyanin deposition throughout plant tissues including anthers, scutellar node, prop roots, roots, stem, ligules, leave, and aleurones via its action as a transcription activator (Bako 2011,

Consonni 1993) (Figure 6). As shown in Figure 6, *r1* activates the *chalcone synthase (CHS), dihydroflavonol 4-reductase,* and *udp-glucose flavinoid 3-oxy-glucisyl transferase* biosynthetic enzyme genes in the anthrocyanin synthesis pathway (as reviewed in Dooner *et al.* 1992*)*. This regulation of the CHS enzyme by *r1* is a function of the r1 protein, which acts as a transcription factor via a DNA binding region of the protein which is homologous to a helix-loop-helix motif of the *myc* family of proto-oncogenes for activation (Dooner *et al.* 1992*,* Ludwig *et al.* 1989).

***R-sc:86-17pale* and its origins:**

The *R-sc:86* allele was recovered by Ashman (1960) as a solid black/purple seed color mutation from *R-stippled* (*R-st)*; which has spotted seeds. *R-st* and *R-sc:86* are both complex alleles; *R-st* has 4 *r1* kernel color genes, *R-sc:86* has 3 *r1* kernel color genes (Table 1, Figure 7). The presence of the *Inhibitor-of-R* (*IR)* Transposable element (TE) in *R-stippled's Sc* gene inhibited expression of the *Sc* gene and results in yellow aleurones, and loss of the *IR* element from *Sc* by *IR* excision or unequal crossover results in black kernels (Eggleston *et al.* 1995). The presence of *Doppia* TEs in the *Nc* genes (Near colorless/lightly mottled) affect the *Nc* expression patterns, due to the 5' end of the *r1* genes controlling tissue specific expression (Figure 7) (Eggleston *et al. 1995*). *R-sc:*86 likely arose from an unequal crossover event between the *Sc* and the *Nc1* which created a chimeric gene (*Sc||nc1*), and has the functional 5' end of the *Sc* gene fused to the functional 3' end of *Nc1*. The presence of the *Doppia* insertions in *Nc1* and *Nc2* causes a significant RNA transcript length difference between the *Nc* and the *Sc||nc1* genes. (W. Eggleston, pers. comm., Mary Alleman, per. comm., R. Okagaki & J.

Kermicle, pers. comm.). Shallow sequence analysis indicates that the *Nc2* and *Nc3* genes are homologous to each other, as well as to the 3' end of the *Sc||nc1* gene, with the similarity beginning in the third exon (W. Eggleston, pers. comm.). The *Sc||nc1, Nc2,* and *Nc3* genes in *R-sc:86-17pale* and in the habituated sublines have been found via Southern blot analysis to contain no insertions or deletions larger than 100 bp relative to *R-sc:86* or to each other (W. Eggleston, pers. comm.) (Figure 10).

In 1961, while screening for mutants of *R-sc* alleles, McWhirter identified a likely pre-meiotic mutation on an ear of *R-sc:86,* where a cluster of pale kernels on the ear were lighter than the normal black kernels on the rest of the ear (McWhirter 1961). From this cluster of pale kernels, McWhirter isolated homozygous lines for each of the pale kernels including *R-sc:86-17pale* (K. McWhirter pers. comm.). McWhirter observed a high level of seed-color variation in homozygous *R-sc:86-17pale* ears as well as the other 6 pale lines. As shown in Figure 8, over five generations, McWhirter was able to alter the *R-sc:86-17pale* line into several semi-stable and heritable colored sublines through self-pollinated selection for lightest and darkest kernel color over successive generations (Figure 8) (K. McWhirter, pers. comm.). These seed color changes, as was demonstrated by Meins and Lutz (1979), occurred at a rate significantly faster than mutation rates can account for, and are also reversible, indicating that the kernel color differences between the *R-sc:86-17pale* sublines are not the result of a series of mutations because if the sublines were the result of individual mutations, they would take an average of twenty generations to change color, and would do so randomly (Maroof *et al.* 1961).

In order to determine if the accumulation of genetic modifiers of seed color (other loci) were affecting the expression levels of the *R-sc:86-17pale* sublines, as has been

found to be possible in *r61* genes by Arteaga-Vazquez *et al.* (2010), mapping studies were performed by W. Eggleston (pers. comm.). In the mapping study, *R-sc:86-17pale* Lightest/*r-r:n-19* and *R-sc:86-17pale* Lightest/*ryw* heterozygotes were test crossed and 17 black seed kernels were recovered at a rate of ~1:1000 (W. Eggleston, pers. comm.). This lack of true-breeding intermediate colors in the mapping test cross progeny is consistent with only a single affector of kernel color, reducing the likelihood of modifiers interacting with the *r1* gene to produce habituation (W. Eggleston, pers. comm.). Southern blot analyses of 11 progeny of the 17 of the recovered black kernels using Restriction Fragment Length Polymorphisms found that all had recombination between the *Sc||nc1* gene of the *R-sc:86-17pale* Lightest and the *P* gene (color gene that does not cause kernel color) of *r-r:n19* near the 3' end in the second intron, resulting in a *Sc||p* chimera as shown in Figure 11 (Derkits 2013) (Figure 11). Sequencing analysis of six of the recovered black kernels narrowed the possible area of recombination between *Sc||nc1* and *P* to a 3,500 bp region starting ~650 bp upstream of the end of the second intron of *Sc||nc1* through comparison of SNPs and sequence polymorphisms (W. Eggleston, pers. comm.)

W. Eggleston (pers. comm.) has also performed crosses to test the effect of miRNA or RNAi on the kernel color. W. Eggleston performed crosses with the *R-sc:86-17pale* Lightest sublines to lines containing a *Mediator of Paramuation 1* (*Mop1*) mutation that disrupts the Dicer pathway, and disrupts production of RNAi, affecting RNAi-mediated regulation (Arteaga-Vazquez and Chandler 2010, Woodhouse 2006). Crosses of the *R-sc:86-17pale* Lightest subline to *Rmr6* (*required to maintain supression 6*) mutant lines which do not have a functional RNA polymerase IV and which also have

the effect of suppressing RNAi regulation and of removing RNAi-based paramutation, did not show a significant change in kernel color in the first generation, but did show the light mottling of *Nc* activity or very light *Sc* expression (W. Eggleston, pers. comm.) (Figure 12). However, the second and third generations of the lightest x *rmr6*- crosses had an increased level of light kernel mottling, which may indicate some role of miRNA control of kernel color regulation or activation of *Doppia*.

Potential role of epigenetics in the origin of*R-sc:86-17pale*:

The ability to change the kernel color of *R-sc:86-17pale* within a few generations, and to reverse the change, is a genetics-based evidence against sequence mutation-based changes being directly responsible for the *R-sc:86-17pale* color sublines (Averitt *et al.* 2017, Banerjee *et al.* 2017, Kusmec *et al.* 2017). The lack of large insertions or deletions, and the changes in methylation patterns indicate that the kernel color phenotypic labiality is not the result of a single mutation in the gene driving changes in kernel color (Figure 10). If a sequence-based mutation, it is not likely to be found in the kernel color gene itself, but more likely occurred at a distance from the gene, as has been reported in similar situations, such as the paramutation found in the similar *B* locus, where changes in expression pattern were affected by genomic changes to an enhancer ~100 kb upstream of the transcription start site (Arteaga-Vazquez & Chandler 2010, Stam *et al.* 2002). A set of 853 bp tandem repeats located ~100 kb upstream appears to be required for *b1* paramutation in maize, but the changes in expression from paramutation do not follow the same pattern as the habituated sublines. Paramutation as defined by Brink (1956) can be either an up-regulation or a down-regulation of a gene,

but does not allow for stepwise changes in expression, or increases and decreases in succession (Le Roy *et al.* 2017, Fisher *et al.* 2017)

W. Eggleston (pers. comm.), using methylation sensitive and insensitive restriction digestion enzymes found changes in CHG, and CpG methylation of all three *r1* genes via Southern blot in *R-sc:86* relative to the *R-sc:86-17pale* habituated sublines relative to each other and to the then published gene standards (Figure 10) (W. Eggleston, pers. comm) (Bewick & Schmitz 2017, West *et al.* 2014, Bewick *et al.* 2017). In Figure 10, all lanes were digested with *Hin*dIII, which has 3 restriction sites in the canonical *r1* gene sequences (Figure 10). With the exception of lane 1, DNA was also digested with either *Bst*N1 (B) (methylation insensitive,) or *Eco*RII (E) (blocked by CHG methylation), which are isoschizomers, to provide a comparison of methylation levels at those sites in the *r1* genes. Shown below the allele names are the number of *r* genes present in each allele, which can be seen to have a corresponding effect on the band intensity for each allele, with *R-st* displaying higher intensity bands (4 *r1* genes) than *R-sc:124*, which has only 1 *r1* gene (Figure 10).

The pattern shown in Figure 10 of decreasing intensity of the 0.95 kb band in methylation sensitive digests in the habituated sublines and a corresponding increase in intensity of the 7.5 kb band indicates that there is an inverse relationship between kernel color and methylation at the 3' end of *r* genes in the habituated sublines. CHG and CG methylation increases as kernel color intensity decreases towards yellow (no expression) (W. Eggleston, pers comm.). This trend in CHG methylation indicates that at the 3' end, the kernel color and the methylation level is inversely related to kernel color, and is significantly different from the progrenitor (W. Eggleston, pers. comm.). This

methylation pattern is supported by not just data from *Eco*RII and *Bst*N1, but also with data from southern blots using *Hae*II (CG methylation sensitive) and *Pvu*II (CHG methylation sensitive) to detect methylation at recognition sites.

Using the *R-sc:86-17pale* line as a baseline, Figure 10 shows that as the kernel color decreases, there is an increase in CHG cytosine methylation at the 3' end of the three *r1* genes in habituated sublines. This pattern of increased methylation at the 3' end of the three *r1* genes is in contrast to the pattern of paramutation-induced silencing at the *r1* locus, which exhibits increased 5' methylation of a silenced or partially silenced gene (Walker 1998, Eichten *et al.* 2011). This pattern of increases in 3` CHG cytosine methylation being inversely correlated with kernel color also differs from the usual gene methylation at the 3` end (the last 1/3 of the gene body, starting near the end of the second intron), where methylation at the 5` end has been found to be correlated with gene expression, rather than inter gene body 3` methylation (Bewick and Schmitz 2017, West *et al.* 2014). This understanding of methylation patterns and expression patterns has since changed, to include not just methylation in the promoter region, but also methylation in the gene body (Bewick *et al.* 2017, Bewick and Schmitz 2017, Springer *et al. 2018*). Southern blot analysis indicates that the methylation patterns indicative of paramutation do not match the methylation patterns of the habituated sublines. The methylation patterns, coupled with test crosses with the lightest and darkest sublines that showed no paramutagenic effect, indicate that paramutation is not the cause of the habituated subline kernel coloration, which will be further tested with sequencing and bisulfite sequencing results (Figure 10).

Relevance:

The study of evolutionary processes has focused on accumulated sequence mutations that result in phenotypic changes, but new research is showing that evolution may also be affected by epigenetic changes, and that different epigenetic gene expression patterns can affect fitness (Bedford *et al.* 2012, Day 2016, Mendizabal *et al.* 2014). Many of these epigenetic changes are conserved across species, and can be studied in terms of population genetics, as Lamarck's theory of 'use and disuse' inheritance postulates (Lamarck 1809, Mendizabal *et al.* 2014, Savickiene *et al.* 2017, Teh *et al.* 2017, Werner *et al.* 2017). Better understanding of the mechanisms of epigenetics and habituation will contribute to a better understanding of population genetics and evolution. Research in crop genetics has largely been based on mendelian genetics, and has produced large increases in crop yield (Figure 13) (Yue 2014). Much of this progress had been made through the use of standard breeding and crop development techniques, but these techniques require a multi-decade timeframe (Maroof *et al.* 2009). This time frame can be shortened through the use of marker assisted selection, and through the use of transgenic techniques, as was performed to create glyphosphate resistant plants, but further methods are needed to reduce the developmental time frame and avoid transgenics' poor public perception (Teh *et al.* 2017, Yeu 2014, Kuchel *et al.* 2005). The potential to speed development of a plant species, either through uncovering cryptic alleles already present in the organism or by changing the expression patterns of existing alleles, requires a better understanding of the mechanisms of epigenetics and habituation (Largo-Gosens *et al*. 2016, Li *et al.* 2005). By gaining a better understanding of the mechanisms behind canalization and habituation, it may become possible to not just

globally deregulate canalization, but to adjust the canalized state of single genes through habituation, which would allow development of new lines that utilize formerly cryptic variation, and of new applications for other organisms, as well as allowing a better understanding of epigenetic effects in population genetics.

Specific Aims:

In order to characterize the molecular changes associated with the kernel color changes of the *R-sc:86-17pale* allele at the maize *red color1* (*r1*) locus to determine if the changes are epigenetic in nature, three significant aims have been completed:

1. In order to quantify the kernel color differences between the *R-sc:86* and habituated sublines and to determine if the color differences conflict with the differences expected from habituation, color quantification was carried out using an Agtron M-45 Colorimeter to asses color density, and the results were statistically analyzed.

2. To determine whether there is a significant sequence difference between *R-sc:86* and the habituated sublines at the 3' end of the *Sc||nc1* gene that could account for changes in seed color, DNA sequence analysis has been performed on *R-sc:86*, the *R-sc:96-17pale* Lightest, and *R-sc:86-17pale* Darkest sublines using Sanger sequencing, and assembled for sequence comparison.

3. In order to compare the methylation levels of the *R-sc:86* and *R-sc:86-17pale* lines at the 3' end of the *Sc||nc1, Nc2,* and the *Nc3* genes, analysis of sodium bisulfite-based cytosine methylation was performed on DNA purified from the *R-sc:86:17pale* Lightest and *R-sc:86-17pale* Darkest habituated sublines and from *R-sc:86* to test if increased methylation in the *Sc||nc1, Nc2,* and the *Nc3* genes is inversely correlated with

kernel color changes. Analysis was performed using the sequences produced from Aim 2, and will look for significant differences between the *R-sc:86,* lightest, and darkest lines.

Chapter Two

Quantificative comparison of seed color of *R-sc:86-17pale* and selection sublines to verify that the kernel color differences are statistically significant

Introduction:

Identification of *R-sc:86-17pale* and recurrent selection over five generations based on seed color differences (light vs. dark) on individual ears was performed by K. McWhirter (pers. comm.) to create sublines with semi-heritable differences in seed color on entire ears (Figure 8) (as described in Chapter 1). These sublines were maintained for color (*R-sc:86-17pale* Lightest, *R-sc:86-17pale* Medium, and *R-sc:86-17pale* Darkest seed color) by W. Eggleston (pers. comm.). Light reflectometry was used to quantify seed color in order to test for statistical differences (Etchevers 1976) between the selected sublines with each other, with the fully colored progenitor *R-sc:86* (Gardner 2003). Results were also compared to *r-g:Δ902* and *R-sc:124* (See Table 1), which served as no color and full color controls, respectively.

Methods:

Generation of samples:

Plants homozygous for *r-g:Δ902, R-sc:124, R-sc:86, R-sc:86-17pale* Light, *R-sc:86-17pale* Medium Light, *R-sc:86-17pale* Medium, *R-sc:86-17pale* Medium Dark, and *R-sc:86-17pale* Darkest sublines were grown in duplicate families of 90 with all 90 kernels from the same ear (duplicate families are multiple sets of 90 kernels from the same ear). All alleles were in the W22 genetic background, which has a uniform seed

color, with structures (except *r-g:Δ902* and *Sc:124*) shown in Figure 7 (W. Eggleston, pers. comm.). Families are defined as specific lineages of maize (such as *R-sc:86*), with subfamilies being lines that underwent kernel color selection and were differentiated from a families (line) into subfamilies, and a sample being the kernels from a shelled ear of maize, used for light reflectometry. Optimal fertilization of the field was carried out by testing soil samples (VT Agronomy lab), and recommended amounts of fertilizer and lime applied prior to planting. Plantings of the duplicate families from the same ear were separated by two to three weeks depending on weather and field conditions (rain, temperature and soil conditions). Plants in each planting were self-pollinated over the fewest days possible based on plant and weather conditions in order to reduce the effect of environmental variation. Mature ears were harvested on the same day, dried with fans under ambient conditions, and stored at room temperature. Kernels were removed from the ears with a hand sheller, and sorted to remove defective, dead, and damaged seed. Glumes, silks, and other debris were discarded.

Sample Quantification:

After cleaning, kernels from each ear with sufficient kernels to cover the bottom of the Agtron M-45 sample cup at a depth of at least two layers (>150) were analyzed with an Agtron M-45 colorimeter to quantify the color density of the kernel samples (**Figure 14**). The M-45 Agtron colorimeter was turned on and allowed to warm up for 1 hour prior to each use, and was recalibrated before each session, as recommended by the manufacturer. Each sample was quantified five times to ensure uniformity, with the sample cup removed from the Agtron after each reading and the kernels poured out of the sample cup to another container, and then placed back into the cup for the next reading.

The sample set from each year was analyzed separately to account for environmentally induced differences in seed color between growing seasons. The five sample measurements for each ear were averaged for each subfamily and the controls and were compared by year to test for color differences between both the sublines and between the sublines, the progenitor line, and the no and full color controls.

Statistical Analysis of Samples:

In order to analyze the results from the color quantification of the kernels from each ear, a mean average was generated from the five reflectometer measurements (from each ear,) and sample measurements with a standard deviation of more than 2 were discarded as not representative of the color of the sample, potentially due to an abnormally high proportion of kernels oriented in the sample cup with the embryos facing the light reflectometer. The mean for each sample was then checked for both a bimodal distribution and a normal distribution within the line or subline that the ear is a part of in the year grown using the Shapiro-Wilk normality test. This test for normality allows the use of ANOVA, which requires a normal distribution. An ANOVA was used to test for statistically significant differences between all lines grown within a year, and in order to reduce the false positive rate from multiple unadjusted testing, a Tukey's Honest Significant Differences test was used to simultaneously compare individual lines for significant differences in or between the year grown (Tukey 1949). This comparison uses an adjusted p-value that shows the probability of similarity between multiple comparisons of colorimeter readings.

Results:

All samples (one sample is one ear) were tested for normal distribution within the individual families using a Shapiro-Wilk test for normality (Table 2). Families 2011-080, 2011-081, and 2014-328 had probability values below the threshold of statistical significance (<0.05), indicating a non-normal distribution. These three families were examined for bimodality, and after normal distribution (with outliers) was found, outliers more than 2 standard deviations from the mean were discarded to allow for further statistical testing with ANOVA and Tukey's HSD test.

2011 samples:

An ANOVA analysis of the six 2011 Families, representing 3 subfamilies (Table 2) comprising 104 ears was performed, and a p-value of $<2\ e^{-16}$ was found, indicating a statistically significant difference between the families grown in 2011 (families planted as shown in table 2). The no color control *r-g:Δ902* was not available for sample collection. This shows that a difference between the families tested was detected, but does not indicate which families are significantly different from each other or from the average reflectance measurements of all of the families.

Analysis using Tukey's Honest Significant Differences test was performed on the 2011 families organized by kernel color, and shows that there are significant differences between all three of the epitypes (kernel phenotypes) harvested in 2011, including between the darkest and the full color control (Table 3) (**Figure 15**).

2012 samples:

ANOVA analysis of the 38 ears from the 4 families grown in 2012 (Table 2) produced a p-value of 0.0008, indicating that there was a significant difference in kernel color between the families grown in 2012. Further analysis using Tukey's HSD showed that the *R-sc:86-17pale* Lightest vs. *R-sc:86-17pale* Darkest, *R-sc:86-17pale* Lightest vs *R-sc:86-17pale* Medium Light, and *R-sc:86-17pale* Medium Light vs. *R-sc:86-17pale* Darkest kernel color families each had significant differences with each other, but no significant difference was detected between the *R-sc:86-17pale* Lightest to *R-sc:86-17pale* Medium Light and *R-sc:86-17pale* Medium Dark subfamilies (Table 4, Figure 16).

2014 samples:

ANOVA analysis of the 81 ears from the 7 families grown in 2014 (Table 2) produced a p-value of $<2\ e^{-13}$, indicating that there was a significant difference between families grown in 2014. Further analysis of the subfamilies and the no-color control using Tukey's HSD showed that there was not a significant difference between the *R-sc:86-17pale* Medium Light and *R-sc:86-17pale* Medium Dark families, nor between the *R-sc:86-17pale* Lightest families and *r-g:*Δ902. Tukey's HSD did show that that there was a significant difference between *R-sc:86-17pale* Lightest vs. *R-sc:86-17pale* Medium Light, *R-sc:86-17pale* Lightest vs. *R-sc:86-17pale* Medium Dark, *R-sc:86-17pale* Lightest vs. *R-sc:86-17pale* Darkest, *R-sc:86-17pale* vs. *R-sc:124*, *R-sc:86-17pale* Medium Light vs. *R-sc:86-17pale* Darkest, *R-sc:86-17pale* Medium Light vs. *R-sc:124*,

R-sc:86-17pale Medium Dark vs. *R-sc:86-17pale* Darkest, and *R-sc:86-17pale* Medium Dark vs. *R-sc:124* (Table 5, Figure 17).

Discussion:

In order to determine whether there was a quantifiable difference between the selected subfamilies based on light reflectometry, analysis was performed to determine if the kernel color differences conflicted with the differences expected from habituation. Because environmental effects on the kernel color vary by year and time of year, direct analyses between families grown in different years not possible as a direct comparison across years. This comparison required full color (*R-sc:124*) and no color control families (*r-g:Δ902*) where possible in each year because both controls did not produce sufficient ears each year, which did not allow normalization of the sample set to deal with year-to-year effects (Table 1). The environmental effects that can affect the kernel color include annual differences in temperature, total light levels, rainfall amounts and timing, soil fertility, fertilizer levels, humidity and growth during plant and seed development and growth (W. Eggleston, pers. comm.). Soil nutrients affect plant growth, and although efforts were taken to ensure that fertilizer was evenly spread, the equipment used did not apply consistent fertilizer throughout the field, which also affect between year variations. Harvesting time also varied from year to year, as a function of weather. This variation in harvesting means that there were different amounts of time for seed maturation and drying on the stalk, which can have a minor effect on both kernel color, and on kernel count (Dos Santos *et al.* 2005).

Comparison of kernel color within years has the effect of minimizing environmental effects, because all of the kernels from that year that were planted at the same time have undergone approximately the same amount of heat, light, and environmental stresses. All samples in each line were tested for normal distribution using the Shapiro-Wilk test, and of the families tested in 2011, 2012, and 2014, only three had non-normal distribution, in all three cases due to single outliers rather than due to multi-modal distribution (Table 2). The appearance of outliers is consistent with the kernel color variations that occur due to the loss of canalization and the habituation of the subfamilies, but is also consistent with the light-inducible nature of *r1* gene, as some of the variation may be due to different ears and parts of ears receiving varying levels of light, which would cause increased levels of kernel pigmentation and variation. This generally normal distribution is also consistent with the color selection process that discourages bimodal distribution through hand selection of the desired kernel color for each subline prior to planting.

The ANOVA analysis of the samples from each of the three years tested showed significant differences between sample families in a given year, with varying levels of significance (2011= $<2\ e^{-16}$, 2012=0.0008, and 2014= $<2\ e^{-13}$), indicating that there are real color differences between the subfamilies and the progenitor line, but not indicating which families are different. In order to examine differences between individual families in each year, Tukey's HSD analysis was performed to reduce the type 1 error rate (increased false positive rate) that would be caused by pairwise ANOVA. Tukey's HSD showed that two of the three of the 2011 families were significantly different from each other, but showed that the 2012 families were not all significantly different from each

other, nor were the 2014 families entirely significantly different from the other families of their respective years. The three 2011 families showed that there was significant differences between all three families, but that the largest difference was between the *R-sc:86-17pale* Lightest and the *R-sc:86-17pale* Darkest of the habituated subfamilies, as expected from visual examination of the subfamilies (Table 3, Figure 8). The 2012 families had significant differences between several of the families, with differences between the *R-sc:86-17pale* Darkest and the *R-sc:86-17pale* Lightest, but not between the *R-sc:86-17pale* Medium Light and *R-sc:86-17pale* Medium Dark families, as can also be seen in the 2014 results. These differences between the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest families are consistent in all three years tested, which is consistent with the visual observable data of light yellow (*R-sc:86-17pale* Lightest) and dark purple (*R-sc:86-17pale* Darkest) being significantly different (Figure 8). This significant difference between the selection sublines indicates that any molecular basis for the kernel color changes is most likely largest between the *R-sc:86-17pale* Darkest and the *R-sc:86-17pale* Lightest sublines, suggesting that looking for molecular differences in sequence between the two extremes of the selection sublines may show any sequence differences.

Light reflectometry did not show a statistically significant difference between the *R-sc:86-17pale* Medium Light and *R-sc:86-17pale* Medium Dark families in the two years that data could be collected for those families. This lack of significant quantitative difference is not consistent with the visual selection of seed that was used to create the two subfamilies. The lack of difference may be partially explained by the light inducible nature of the kernel color expression, which may introduce a difference in the expression

of the kernel color in the aleurone and the expression level the embryo will produce in the next generation. This could indicate that although light reflectometry can provide qualitative data, it is somewhat different from visual selection for a shade of color, due to the visual selection being based on crown color, while the light reflectometer quantifies light reflectance from all parts of the kernel in an aggregate sample from the entire ear. This reliance of light reflectometry on the aggregate reflectance may not detect the visual differences between the kernel crown colors, as the color differences may be lost in the aggregate color of the rest of the kernel (embryo, sides, crown.)

As shown in the comparison of the *R-sc:86-17pale* Lightest and the *R-sc:86-17pale* Darkest subfamilies, and the methods used to make the subfamilies show, the changes in kernel color are happening more consistently, and at a rate that is far faster than that which is hypothesized by Mendelian genetics (Figure 7, Figure 8). In Mendelian genetics, a mutation rate of approximately 8 in 10,000 is expected, which is a far smaller rate of change than is found in the subfamilies (Stadler 1946) (Figure 8). In addition to the inconsistency in rate of color change between the subfamilies and the rate of change posited to be Mendelian, the reversibility of the kernel color changes also argue against the change being entirely due to Mendelian mutation (Figure 8). The rate of the color changes, and the reversibility of those changes, is more similar to the changes found in habituation, an epigenetic phenomena (Meins & Lutz 1979, Meins 1989). Habituation, or the change in a characteristic over time in response to external stimuli, often happens much faster than can be expected from mutation-based changes, and is not the result of individual genomic mutations driving each phenotypic stepwise change (Meins & Lutz 1979, Largo-Gosens *et al.* 2016, de Castro *et al.* 2017). This pattern of

rapid change over a small number of generations is very similar to that seen in the kernel color changes shown in Figure 8, and is consistent with the idea of a mutation event causing the original color shift in *R-sc:86-17pale*, not by direct mutation, but rather by the mutation affecting the mechanisms of control of kernel color or an epimutation as described in Chapter 1 (Gage *et al.* 2017). The idea of the cause of kernel color plasticity being epigenetic in nature is supported by southern blot work that shows that there are epigenetic differences between the lightest, darkest, and parent families in the *Sc||nc1* gene, as described in Chapter 1 (Figure 10, Table 1). The hypothesis that the selected kernel color variation/change is not based upon individual mutations, but rather upon epigenetic causes.

Chapter Three

DNA sequence comparison at the 3` end of the *Sc*||*nc1* gene between the *R-sc:86-17pale* Lightest, *R-sc:86-17pale* Darkest, and *R-sc:86* lines

Introduction:

As found in Chapter 2, the kernel color of the Lightest and the Darkest *R-sc:86*-17pale sublines are significantly different from each other. Two primary hypothesis can account for these differences, ranging from pure genetic to epigenetic based hypothesis. One of these hypothesis is that the kernel color changes are caused by a series of sequential mutations which effect kernel color. Another hypothesis is that the kernel color is the result of a mutation that is allowing epigenetic effects on kernel color to be observed and altered via selection. Neo-Mendelian genetics proposes that each phenotypic change is the result of a distinct mutation, and an occurrence of 8 mutations per 10,000 kernels is the expected rate of random mutations in the kernel color genes changing kernel color (Stadler 1946). The gene-specific (genic) mutation rate in corn is ~1.32 per 100,000 kernels (Clark 2005, Haberer *et al.* 2005), and therefore, a phenotypic trait-specific change is expected to take multiple generations to be expressed, isolated, and stabilized in a population and is not quickly reversible due to the mutation needing to be reversed by another mutation (Maroof *et al.* 2009). The hypothesis of a series of sequential sequence mutations at each step of the phenotypic change having a molecular basis of a distinct mutation has been well documented, but does not fit the results found with the *R-sc:86* line and the *R-sc:86-17pale* sublines (Ashman 1960, Chen 1992, Maroof *et al.* 2009, Averitt *et al.* 2017) (W. Eggleston, pers. com.). McWhirter's original selection work using self pollination and selection of multiple light and dark kernels per ear was

based on many kernels diverging in color on each ear, rather than single kernels each generation, and the ability to reverse the kernel color change within the same number of generations (Figure 8) (W. Eggleston, pers. com.). The ability to affect phenotypic change (and to reverse it) at the rate of change demonstrated by the *R-sc:86-17pale* sublines, suggests that the changes in kernel phenotype are not the direct result of multiple separate Mendelian mutations causing each phenotypic change, but rather that the kernel color changes have another molecular basis, as the rate of change is significantly higher and is much more consistent than a mutation-based system would be expected to allow (Figure 8). In order to better understand the molecular basis(es) of the progressive kernel color changes, Eggleston (pers. com.) performed test crosses of *R-sc:86-17pale* Lightest heterozygous with two different *r1* plant color alleles with only one *r1* gene that activated plant color (*P* gene), but not kernel color (*r-r:n19* and *ryw*) (Figure 11). Test cross progeny were screened for kernel color changes from the yellow kernels conditioned by *R-sc:86-17* Lightest, *r-r:n19* and *ryw*. Out of approximately 78,000 kernels screened, 17 revertants to full kernel color (*R-sc:86* phenotype) were recovered. No bona Fide intermediate colored kernels were recovered in the screen. Restriction enzyme map analysis showed that all 11 revertants tested by Southern blot analysis had a recombination event in or just beyond the 3` end of the *Sc||nc1* gene (Derkits 2013) (W. Eggleston, pers. com.). The location of these recombination events in all of the revertants tested indicates that the molecular basis for the phenotypic change is located in the 3` end of, or 3` to the *Sc||nc1* gene 8 (Derkits 2013). These recombination events provide a starting place for sequence analysis of the kernel color variation, both to

attempt to identify the molecular basis and to determine if different mutations at the 3` end of the *Sc||nc1* gene are affecting the kernel color changes.

Methods:

DNA was extracted from crown tissue samples from families from the *R-sc:86-17pale* Lightest, *R-sc:86-17pale* Darkest sublines, and from *R-sc:86* families as detailed in Lindsay and Eggleston (2014). DNA from the original *R-sc:86-17pale* was not available. DNA from the 3` end of the *Sc||nc1* gene was amplified with Longamp *Taq* DNA polymerase using the Sc7737 and Lc8701 primers to provide specificity to the *Sc||nc1* gene (Figure 9). Amplification was performed by placing the Sc7737 forward primer in the *Sc* region of the *Sc||nc1 gene* in a Hybaid Thermal Reactor (Model HBTR1, Thermo Hybaid, Middlesex, UK), with the following reagents, and under the following conditions: 10 mM Tris-HCl, 50 mM KCl, 1.15 mM $MgCl_2$, 0.1 mM dNTPs, 10 ng/ul forward primer, 10 ng/ul reverse primer, 100,000 Units/ul *Taq*, 20 ng/ul Genomic DNA in a total volume of 25 ul. 94°C for 2 minutes, followed by 35 cycles of 94°C for 1 minute, 55°C for 1 minute, 68°C for 5 minutes (Figure 9). PCR reactions were carried out under mineral oil, and the *Taq* DNA polymerase was added in a Hot Start fashion, after the first denaturing step had finished (Figure 9). Following amplification, the amplicon was treated with exoSAP to de-phosphorylate primers and dNTPs and then sequenced at Yale's Keck DNA sequencing facility using the primers listed in Figure 9. Sequencing took place on both backbones using both forward and reverse primers, with a primer spacing of approximately 300 bp, and a target coverage of 3x to 5x to reduce the effect of sequencing errors, such as those caused by *Taq*. Once received from Yale, raw

sequencing reads were trimmed and based called using the UGene sequence analysis program (Okonechinikov *et al.* 2012). The sequences were then assembled using the CLC 7 sequence assembly program. Sequence contigs for the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines and *R-sc:86* were assembled separately with the goal of 5x coverage on each backbone. Once contigs were assembled for the *R-sc:86-17pale* Lightest, *R-sc:86-17pale* Darkest, and *R-sc:86-17*, a consensus sequence was created for each assembly. The consensus sequences were aligned in the CLC7 sequence assembly program and analyzed for sequence homology.

Results:

The analysis for homology between the lines from the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines, and the *R-sc:86* at the 3` end of the *Sc||nc1* gene showed a consensus sequence with six potential sequence differences in a region spanning 2,970 bases. Within the consensus of the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines, and the *R-sc:86* there is an ~12 bp poly T region located in intron 4, starting at base 8,545 of the reference *R-Sc124* alignment sequence (M. Alleman & W. Eggleston, pers. comm.). Sequence coverage near this region did not reach the desired 5x coverage. Sequence reads extending past the poly T region on the top strand ranged from 2-4x coverage until the next primer location. Sequence on the bottom strand extending upstream of the poly T region was missing until the next primer location (**Figure 18**).

Of the six identified potential sequence differences identified when compared to the *R-Sc:124* reference sequence, five were in the *R-sc:86-17pale* Darkest subline, and the other divergence was in the *R-sc:86-17pale* Lightest subline (

Table **6**). Sequence analysis indicated that only 2 of the 6 sequence potential differences were located in a coding region, and of the single nucleotide differences found, all were the result of ambiguity in the sequencing results, rather than a distinct difference in sequence from the other phenotypes which would cause a change in the coded amino acid. None of the potential differences were indels indicating no change in sequence length or protein sequence length between the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines, or the *R-sc:86* line. No confirmed unambiguous sequence differences were identified.

Discussion:

Sequencing of the 3` end of the *Sc||nc1* gene was performed to test for sequence differences between the *R-sc:86* line and the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines relative to each other and to *R-sc:86*. This sequencing was carried out on the *R-sc:86* line, and was carried out on the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines in order to test the extremes of the kernel color changes. Were sequence difference to be found, the stepwise progression of kernel color change sublines detailed in Chapter 2 would be tested for sequence changes. Sequence homology between the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines, and the *R-sc:86* line indicates that there are no unambiguous changes in DNA sequence at the 3` end of the *Sc||nc1* gene, and that there is no amino acid differences found in the region of the *Sc||nc1* gene sequence between the tested *R-sc:86-17pale* sublines and the *R-sc:86* line. These possible DNA sequence differences are single base differences rather than indels and are not known to affect protein length or RNA transcript length via stop

codons, and would not result in a frame shift or changes to more than one amino acid in the *Sc||nc1* protein. Common to all of the sequences is a 12-13 bp poly-T region, the length of which is ambiguous (**Figure 18**, **Figure 19**, **Figure 20**). Of the 6 individual possible DNA sequence differences in sequence, 4 of the 6 possible differences were found in introns, and would not affect the amino acid sequence (

Table **6**). These intron based possible DNA sequence differences were found in the *R-sc:86-17pale* Darkest subline, and 3 of the 4 possible differences were differences from a consensus sequence of C, possibly due to a bias in either the PCR amplification due to the *Taq* polymerase used, or in the Sanger sequencing technique used in high GC regions such as that sequenced (Chen *et al.* 2013). These possible differences were not clear cut base calls that were different from the base sequence, but rather were sequence differences where all of the electropherograms did not agree as to the base pair in question. Of the two possible DNA sequence differences in exons, both differences are N base calls, rather than base changes from the overall consensus sequence. These base calls are the result of unclear peak structures in the electropherograms from the Big Dye Sanger sequencing, and show signs of the bias, likely due to an elevated GC content (Ross *et al.* 2013). Because both of the intron differences were called as N, the overall effect of the possible differences in the sequence cannot be known, but as the darkest and the *R-sc:86-17pale* Lightest sublines contain an N in different introns, it is not likely that the two Ns conceal kernel color changing mutations in their respective habituated sublines.

In all three lines the size of the poly T region is ambiguous possibly due to polymerase slippage. The poly T region is estimated to be 12 or 13 bases, but the number

of thymine bases is an estimate based on previous sequencing work (W. Eggleston, pers. com.). The poly T region is similar to the poly A tail found on mRNA which leads to transcription termination via polymerase disengagement (Kielaczawa 2006, Mischo & Proudfoot 2013).

Overall, the homology of these three DNA sequences and the lack of un-ambiguous differences in the coding regions, as well as a lack of evidence of DNA sequence changes that could change expression patterns provide strong evidence that the kernel color variation shown in chapter 2 is not the result of a DNA sequence difference in the 3` end of the *Sc||nc1* gene, but suggests another control of kernel color variation, with the molecular basis found in the 3` end of the *Sc||nc1* gene (Figure 10). Previous work done with Southern blots and a combination of methylation sensitive and insensitive isoschizomers indicates that although not necessarily a cause of the kernel color changes, differences in methylation are present at the 3` end of the *Sc||Nc1* gene, and may yield more in-depth information if examined with sodium bisulfite sequencing.

Chapter 4

Analysis and comparison of DNA methylation patterns at the 3` end of the *Sc||nc1* gene between the *R-sc:86-17pale* Lightest, *R-sc:86-17pale* Darkest, and *R-sc:86* lines

Introduction:

As shown in Chapters 2 and 3, the kernel color of the *R-sc:86-17pale* sublines are significantly different from each other, and have no un-ambigous DNA sequence differences at the 3` end of the *Sc||nc1* gene. W. Eggleston(pers. comm.) showed that recombination in the 3` end of the *Sc||nc1* gene causes a reversion in the kernel color, indicating that the 3` end of the *Sc\\nc1* gene is important for expression of the kernel color. Southern blot analysis using a combination of methylation sensitive and insensitive restriction enzymes indicates the there is a methylation difference between the *R-sc:86* line and the *R-sc:86-17pale* selection sublines at the 3' end of the *Sc||nc1* and *Nc2, and Nc3* genes (Figure 10) (W. Eggleston, pers. comm.). Methylation in the *R-sc:86-17pale* Darkest and *R-sc:86-17pale* Lightest sublines was examined to determine if the extremes of kernel color showed methylation differences prior to an examination of the stepwise progression for methylation differences in all of the different kernel color changes in the *R-sc:86-17pale* selection sublines. This was examined in more detail using sodium bisulfite based conversion of genomic DNA from kernel crowns, followed by DNA sequencing analysis of the gene region in question in order to determine the extent of the methylation differences between the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines, and to test for methylation differences, and if found whetehr they are correlated with the kernel color changes in the *R-sc:86-17pale* sublines.

Methods:

Genomic DNA for the *R-sc:86* parent line, the *R-sc:86-17pale* Lightest and the *R-sc:86-17pale* Darkest sublines was extracted from kernel crowns as described in Chapter 3, from maize plants grown as described in Chapter 2 (Lindsay and Eggleston 2014). Two hundred and thirty eight samples from 2014 were treated with sodium bisulfite to convert unmethylated cytosines to thymines using the QiagenEpitect kit, of which five samples were successfully converted (Gaithersburg, MD). The sodium bisulfite-treated DNAs were used to perform nested PCR, with degenerate primers as shown in Figure 21 and Table 7. Nested PCR was performed with 21 different combinations of outer and then inner nests to allow for amplification of the sodium bisulfite treated DNA, of which 3 combinations worked (Figure 21). All outer nests had a forward primer in the *Sc*-specific region of the *Sc*||*nc1* gene, to ensure that the PCR product came from *Sc*||*nc1,* instead of from *Sc**nc1, Nc2,* and *Nc3* simultaneously. The initial successfulouter nest was performed with primers Bs47666M and Bs2Lc6237r, under the following conditions, 10 mM Tris-HCl, 50 mM KCl, 1.15 mM $MgCl_2$, 0.1 mM dNTPs, 10 ng/ul forward primer, 10 ng/ul reverse primer, 100,000 Units/ul *Taq*, 20 ng/ul Genomic DNA, with a total volume of 25 ul in a Biorad T100 thermal cycler (Irvine, CA); denaturation for 30 seconds at 95°C, followed by 40 cycles of 95°C for 30 seconds, 53°C for 45 seconds, and 72°C for 1 minute, followed by a 10 minute final elongation at 72°C. One microliter of the outer nest PCR reaction was used as starting material for the successful inner nest, which was performed with primers Bs47666m and Bs2Sc8045r under the following conditions, 10 mM Tris-HCl, 50 mM KCl, 1.15 mM $MgCl_2$, 0.1 mM dNTPs, 10 ng/ul

forward primer, 10 ng/ul reverse primer, 100,000 Units/ul *Taq*, with a total volume of 25 ul in a Biorad T100 thermal cycler (Irvine, CA); denaturation for 30 seconds at 95°C, followed by 40 cycles of 95°C for 30 seconds, 69°C for 45 seconds, and 72°C for 1 minute, followed by a 10 minute final elongation at 72°C. Amplicons were size fractionated using gel electrophoresis as described in Lindsay and Eggleton (2014), and were cut out of 2% Low Melting Point Agarose gel for purification using the Qiagen Qiaquick Gel Purification kit (Gaithersburg, MD). The purified amplicon was sent to Yale's Keck DNA Sequencing lab for sequencing as in Chapter 3. A later attempt to amplify the sodium bisulfite treated DNA used a different set of parameters and primers. The outer nest used the Bs47527 and Bs2Sc8144r primers under the following conditions, 10 mM Tris-HCl, 50 mM KCl, 1.15 mM $MgCl_2$, 0.1 mM dNTPs, 10 ng/ul forward primer, 10 ng/ul reverse primer, 100,000 Units/ul *Taq*, 20 ng/ul Genomic DNA, and a total volume of 25 ul in a Biorad T100 thermal cycler (Irving, CA); denaturation for 30 seconds at 95°C, followed by 40 cycles of 95°C for 30 seconds, 50°C for 45 seconds, and 72°C for 1 minute, followed by a ten minute final elongation at 72°C. One microliter of the outer nest PCR reaction was used as starting material for the inner nest, which was carried out with primers Bs47666m and Bs2Sc8045r under the following conditions, 10 mM Tris-HCl, 50 mM KCl, 1.15 mM $MgCl_2$, 0.1 mM dNTPs, 10 ng/ul forward primer, 10 ng/ul reverse primer, 100,000 Units/ul *Taq*, with a total volume of 25 ul in a Biorad T100 thermal cycler (Irving, CA); denaturation for 30 seconds at 95°C, followed by 40 cycles of 95°C for 30 seconds, 65°C for 45 seconds, and 72°C for 1 minute, followed by a 10 minute final elongation at 72°C. The amplicon was purified

with the Qiaquick Gel Extraction kit as before, and the purified amplicon was sent to Yale's Keck DNA Sequencing lab for sequencing (Gaithersburg, MD).

Sequences were first trimmed using the UGene sequence analysis program, and then analyzed by aligning the sequences to the reference sequence in the CLC 7 program as in Chapter 3 (Okonechinikov *et al.* 2012). The guanine sites in the reference sequence were tabulated, and using a common reference point in the sequences, the electropherographs were examined for full or partial methylation at each potential methylation site in UGene. Methylation status was determined by measuring the adenine and guanine peak heights with a ruler on the screen in each electropherograph after raising the highest peak to a height of 10 cm, and percentage methylation was determined by the relative ratio of peak heights, with the basal peak noise removed where appropriate by subtracting the basil peak noise height from the Adenine and Guanine peak heights.

Results:

The analysis for cytosine methylation of the *R-sc:86, R-sc:86-17pale* Lightest, and the *R-sc:86-17pale* Darkest sublines has 74 DNA cytosines in the region sequenced, with 42 sites on the top strand, and 32 on the bottom strand; three CG sites, four CHG sites, and 25 CHH sites (H can be Adenine, Cytosine, or Thymine) were present on the bottom strand. Eleven CG sites, 22 CHH sites, and 9 CHG sites were present on the top strand. Of the sites present on the bottom strand, 20 of the 32 cytosine sites were fully methylated in all samples tested, with the remaining 12 cytosine sites being either partially or not methylated (

Table **8**). The fully methylated cytosine sites are two of the three CG sites, 14 of the 25 CHH sites, and all four of the CHG methylation sites (**Error! Reference source not found.**). The third CG site is fully unmethylated, as are two of the CHH sites. The remaining 11 CHH sites are partially unmethylated, with most sites being at approximately the same levels of methylation across samples tested. Four sites had different methylation in the different samples tested. The *R-sc:86* sample had three methylation sites (7759, 7823, & 7835) that showed a difference in methylation between the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines. Methylation sites 7759 and 7835 are fully unmethylated in the *R-sc:86* sequence, but are fully methylated in the *R-sc:86-17pale* sublines. Methylation site 7823 is completely unmethylated in the *R-sc:86* line, but partially methylated in the *R-sc:86*-17*pale* sublines. Two other methylation sites that show differences between the *R-sc:86-17pale* Lightest and Darkest *R-sc:86-17pale* sublines are at position 7787 and 7886. Methylation site 7787 is partially methylated in one of the *R-sc:86-17pale* Lightest lines, but has full methylation in the *R-sc:86-17pale* Darkest subline and the *R-sc:86* lines. Methylation site 7886 shows 10% methylation in the *R-sc:86-17pale* Darkest line, but has full methylation in the *R-sc:86-17pale* Lightest line.

Methylation analysis of the *R-sc:86, R-sc:86-17pale* Lightest, and the *R-sc:86-17pale* Darkest sublines on the top strand had 42 possible methylation sites, and all 42 of the possible sites in all three of the unique sequence contexts show full methylation, with no variation between sequences or between the *R-sc:86* line*,* the *R-sc:86-17pale* Lightest subline, and the *R-sc:86-17pale* Darkest subline.

Discussion:

As was shown by the previous Southern blot work with methylation sensitive restriction enzymes, sodium bisulfite sequencing has shown that there are differences in methylation status between the *R-sc:86* line and the *R-sc:86-17pale* Lightest and Darkest sublines (W. Eggleston, pers. com.) (Table 1) (Figure 10). There are insufficient sodium bisulfite treated samples for the methylation differences to be statistically analyzed, but the evidence is consistent with the evidence that there are differences in methylation patterns of the bottom DNA strand between the *R-sc:86* line and the *R-sc:86-17pale* Lightest and Darkest sublines, and that there are methylation differences between the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines. *R-sc:86* and the *R-sc:86-17pale* Lightest and the *R-sc:86-17pale* Darkest sublines have at least two differences between each, as well as a difference in methylation at site 7886, which is directly adjacent to a *Eco*RII/*Bst*N1 recognition site, and is the base downstream of the cleavage site (Figure 21). The *R-sc:86-17pale* Lightest subline and the *R-sc:86* line are both fully methylated at the 7886 site, while the *R-sc:86-17pale* Darkest subline is ~10% methylated at the site, which may have an effect on the cutting efficiency of the methylation sensitive restriction enzyme *Eco*RII when compared to the fully methylated *R-sc:86-17pale* Lightest subline, which may help explain the pattern shown by the Southern blots of more methylation being present in the *R-sc:86-17pale* subline than in the *R-sc:86-17pale* Lightest line (Table 1) (Figure 10).

While there are individual methylation sites that are consistent with methylation differences between *R-sc:86-17* and the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines, the overall pattern of methylation suggested by the data differs from the published methylation patterns. Bewick and Schmitz (2017) have shown that the level of methylation in the gene bodies in most plants is relatively low, and this pattern of low intra-genic methylation has also been reported in work on maize in a W22 genetic background (Springer *et al.* 2018, Han *et al.* 2018). The pattern of low methylation in the gene body is not uniform across the three types of methylation contexts examined. Springer *et al.* (2018) found that the CG methylation levels in genes are at about 50% methylated, while intragenic CHG methylation sites are about 20% methylated, and CHH methylation was a uniform 10% methylation in the gene body. This published data stands in contrast to the data collected for the *R-sc:86* and the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines, as the methylation levels for CG, CHG, and CHH appear to be much higher than the published results (Table 1,

Table **8**) (Hsu 2017). The CG methylation levels in the three sites present in the sequence are at 90% methylation, while the four CHG sites are all fully methylated instead of the expected methylation rate of 20% (Table 9). The 25 CHH sites are ~90% methylated, which is higher than the expected methylation rate of 10%.

In contrast, the all of the sites on the top strand are fully methylated in the *R-sc:86* line, the *R-sc:86-17pale* Lightest subline, and the *R-sc:86-17pale* Darkest subline. The top strand is not transcribed, and the complete methylation is in contrast to the published results, which do not show a difference between methylation on different strands. This contrast in methylation status between the lines tested on the top strand, and the

published data indicates that the complete methylation is not common, and may be worth further investigation.

These methylation differences are not conclusive for a number of reasons, including small sample size and possible primer induced amplification bias, but the results do provide enough evidence to encourage further study.

Chapter 5

Conclusions on quantifying the kernel color changes, sequencing the 3` end of the *Sc||nc1* kernel color gene, and sodium bisulfite-based sequencing of a region in the 3` end of the *Sc||nc1* gene

Introduction:

In order to determine if the kernel color changes in the *R-sc:86-17pale* selection sublines are epigenetic in nature, characterization of the kernel color of *R-sc:86* and the *R-sc:86-17pale* selection sublines, DNA sequencing analysis of the 3` end of the *Sc||nc1* gene for comparison between *R-sc:86,* the *R-sc:86-7pale* Lightest, and *R-sc:86-17pale* Darkest sublines, and the characterization of DNA cytosine methylation in a part of the same sequences has focused on better understanding the molecular changes associated with the kernel color changes seen in the *R-sc:86-17* selection sublines (Figure 26).

Quantitative analysis of the kernel color changes in the *R-sc:86* line and in *R-sc:86-17pale* selection sublines

Characterization of the kernel color changes show that there are significant differences in kernel color between the *Rsc:86-17pale* selection sublines, *R-sc:86*, and the full-color and no-color controls. These significant differences indicates that the visually different colors in the *R-sc:86-17pale* selection sublines can be quantified and shown to be quantitative, rather than qualitative differences. While the method of quantification used, light reflectometry, is the published standard for quantification of kernel color, there were some challenges with using light reflectometry to distinguish the color density of some of the *R-sc:86-17pale* selection sublines that could be distinguished visually (Stadler 1946, dos Santos 2005). Light reflectometry was not able to distinguish

the difference between four of the selection sublines, which may be attributable to several different causes (Figure 15, Figure 16, Figure 17). One of the likely causes is the effect of the heat on the kernel color and on the pericarp color (W. Eggleston, pers. comm.). Another factor which contributes to the problems with quantification of kernel color is the kernel shape. When kernels develop in a temperate climate, they are often shaped more uniformly, and more of the top and upper sides of the kernels are read by the light reflectometer (Monjardino *et al.* 2006). However, when kernels develop in a warmer climate, they can develop a more irregular shape, which can change the proportion of kernel tissue types the reflectometer reads. (Figure 14). The kernel shape is also heavily affected by the completeness of pollination, which can be adversely affected by excessive heat during pollination (>95°C). A third potential cause of the lack of significant quantifiable kernel color difference is that the visual inspection of the ears is usually carried out with unshelled ears, which only shows the crown tissue, rather than the all around view of tissue types that is reflected in the sample cup for light reflectometry. A fourth cause of the lack of significant quantifiable kernel color difference between visual inspection and the quantification of light reflectometry may be the nature of light reflectometry, wherein the reflectometer can quantify color density, but cannot differentiate between samples with a similar color, but have visually discernable color differences. This may be observed in comparing kernels from the *R-sc:86-17pale* Medium light and *R-sc:86-17pale* Medium dark sublines (Figure 15, Figure 16, Figure 17). Light reflectometry is a significantly more reproducible and bias free method of producing quantifiable results than alternative methods, but is less sensitive than eye in distinguishing mid range colors from each other. Light reflectometry yields useable

results that indicate that there are significant differences between the *R-sc:86-17pale* selection sublines.

DNA sequencing of the 3` end of the *Sc||nc1* gene in the *R-sc:86* line, the *R-sc:86-17pale* Lightest, and the *R-sc:86-17pale* Darkest sublines

DNA sequence analysis of the 3` end of the *Sc||nc1* gene has shown that the sequence is indistinguishable between the *R-sc:86* line, *R-sc:86-17pale* Lightest, and *R-sc:86-17pale* Darkest sublines at the depth of sequence analysis completed in the 3` end of the *Sc||nc1* gene suggested by recombination studies to be associated with the kernel color change (W. Eggleston, pers. comm.), indicating that the kernel color change is not the result of sequential DNA sequence mutations in the 3` end of the *Sc||nc1* gene. This region was identified as being a region where recombination in *R-sc:86-17pale* selection sublines could cause kernel color reversion to the *R-sc:86* black kernel color, and was therefore a region presumed to be involved in the molecular basis of the kernel color change (W. Eggleston, pers. comm.) (Dietrich 1993, Okagaki *et al.* 2018). Suppression of recombination by high levels of methylation in this area and in the highly repetitive flanking regions could have had some effect on the frequency of the recombination events as noted by Okagaki *et al.* (2018), but sufficient kernels with reversion to the *R-sc:86* kernel color were found to map the recombination points to in or just beyond the 3` end of the *Sc||nc1* gene (W. Eggleston, pers. comm). The *R-sc:86* sequence and the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sequences do not have 5X coverage in the entire region sequenced, but there is enough coverage to show that there is no unambiguous sequence difference between the *R-sc:86* line and the *R-sc:86-17pale*

Lightest and *R-sc:86-17pale* Darkest sublines. This lack of unambiguous differences between the sequences analyzed suggests that the molecular basis for the kernel color changes, which was localized to in or just beyond the 3` end or beyond the 3` end of the *Sc||nc1* gene sequenced, is epigenetic in nature, rather than being the result of sequence changes.

Analysis of DNA cytosine methylation sites in a region of the 3` end of the *Sc||nc1* gene

The work performed by W. Eggleston (pers. comm.) with methylation sensistive restriction enzymes showed that there are DNA cytosine methylation differences (CG and CHG) between the *R-sc:86* line and the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines in the 3` end of the *Sc||nc1*, *Nc2,* and *Nc3* in both diploid juvenile plant tissue and in triploid immature crown tissue. Specifically, there is an inverse correlation between kernel color and DNA cytosine methylation at the 3` end of *Sc||nc1, Nc2,* and *Nc3*, which stands in contrast to the published work that describes the more common 5` DNA cytosine methylation of a gene being inversely correlated with gene expression, and does not show methylation differences at the 3' end of genes (West *et al.* 2014, Bewick and Schmitz 2017). Similar DNA cytosine methylation differences in the 5` end of genes have been shown to be correlated with differences in protein expression levels in maize, however, differences in mehtylation at the 5` end of *Sc||nc1* was very low in all *Rsc:86-17pale* sublines tested, and did not correlate with kernel color (Turco *et al.* 2017). The published data suggests that the differences in DNA cytosine methylation levels found between the *R-sc:86* line and the *R-sc:86-17-pale* Lightest and *R-sc:86-17pale* Darkest

sublines at the 3` end of the *Sc*||*nc1* gene may be associated with the kernel color differences (Turco *et al.* 2017). The Southern blot analyse provides a gene wide view of the DNA cytosine methylation status in the *Sc*||*nc1, Nc2,* and *Nc3* genes, but it is limited by the availability of methylation sensitive restriction enzymes with different recognition sites in the 3` end of the *Sc*||*nc1* gene, which are present at a low density (Figure 10, Table 1) (Helentjaris 1986). To obtain more detailed DNA cytosine methylation data from aleurone tissue where the *Sc*||*nc1* gene is expressed, sodium bisulfite treatment of the genomic DNA from crown tissue samples, followed by PCR amplification and sequencing was used to analyze base wise DNA cytosine methylation status for the region sequenced (Hsu 2013). The sodium bisulfite-based sequencing provided some DNA cytosine methylation information, but not enough to draw statistically significant conclusions about DNA cytosine methylation levels or differences, due to a low number of working samples and the short region sequenced. This low number of working samples may be due to either incomplete conversion or tissue samples that were insufficiently enriched for aluerone tissue, and does not have sufficient coverage (3x-5x) to reduce the chance of *Taq* polymerase errors causing differences in sequence. However, the sodium bisulfite sequencing results that were collected do support and extend the Southern blot data which shows DNA cytosine methylation differences between the *R-sc:86* line and the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines (Figure 10,

Table **8**). In the region sequenced, there is a *Eco*RII/*Bst*N1 recognition site, but the site (7884,

Table **8**) shows no difference in DNA cytosine methylation between the *R-sc:86* line, the *R-sc:86-17pale* Lightest subline, and the *R-sc:86-17pale* Darkest sublines. However, as can be seen at site 7886, there is a potential difference in DNA cytosine methylation status between the *R-sc:86-17pale* Darkest subline and the *R-sc:86-17pale* Lightest subline, as well as the *R-sc:86* line (

Table **8**). New England Bio's published data on the DNA cytosine methylation sensitivity of the *Eco*RII restriction site states that the restriction enzyme is sensitive to DNA cytosine methylation within the restriction site, but had no information as to DNA cytosine methylation adjacent to the restriction site affecting the restriction enzyme, despite the enzyme cutting at the edges of the recognition site. This possible difference between the DNA cytosine methylation levels at site 7886 between the *R-sc:86-17pale* Darkest subline and the *R-sc:86-17pale* Lightest subline, as well as the *R-sc:86* line may be affecting the restriction digestion efficiency and could explain part of the difference in DNA cytosine methylation shown by Southern blotting (

Table **8**) (Figure 10). The DNA cytosine methylation results suggest that the DNA cytosine methylation levels in the bottom strand of the *R-sc:86* line and the *R-sc:86-17pale* Lightest and *R-sc:86-17pale* Darkest sublines may be much higher than the published works have shown, on both the tranxcription (coding) strand, and on the complementary strand (Bewick and Schmitz 2017, Springer 2018). The DNA cytosine methylation data also suggests that a higher level of tissue-specific CHH methylation (~70%) is present in the coding strand, and full methylation is present at all sites on the complementary strand, rather than the 10-30% CHH methylation level that is expected in the gene body (Han *et al.* 2018). The CHG methylation sites tested are fully methylated,

which is in contrast to the expected methylation levels of ~20% CHG methylation in the gene body (Springer *et al.* 2018, Eichten 2011). The three CG methylation sites present are ~70% methylated, differing from the ~50% methylation rate found by Bewick and Schmitz (2017). The complete methylation of the 11 CG site, 22 CHH sites, and 9 CHG sites present in the top strand also differs from the published data, which does not differentiate between methylation of the transcribed strand and the complementary (top) strand. This difference is not minor, and is worth further invetigation. The CG and CHG methylation frequencies found here differs from the published rates, but due to the very small number of sites samples tested, the differences in methylation rates found here may be due to sampling bias of the sites, the small number of samples, or the increased sample noise in bisulfite sequencing (Ross 2013, Mach 2013, Hsu 2013) Some of the difference in the overall methylation levels may also be due to the method used for measuring the partial methylation, which by using a peak height ratio is not as inherently accurate as a measurement system based on volume under the peaks. CHH methylation in the *R-sc:86* line, the *R-sc:86-17pale* Lightest, and the *R-sc:86-17pale* Darkest subline has 27 sites, but with the small number of total sequences available, the methylation data is merely suggestive of differences from the published data. This difference in CHH, CG, and CHG methylation rates when compared to the published methylation data suggests that differences in the genic methylation rate and the complementary strand methylation may play a role in expression of the *Sc||nc1* gene, and should be further studied (Han *et al.* 2018, Eichten *et al. ,* 2011).

Examining possible molecular bases of the kernel color change in the *R-sc:86-17pale* selection sublines

Despite the problems encountered, there is sufficient data to examine several different hypothesis regarding the molecular basis of the kernel color changes in the *R-sc:86-17pale* selection lines. The initial hypothesis for a change in phenotypic expression from *R-sc:*86 to *R-sc:86-17pale* expression was a sequence mutation, which most likely was the cause of the initial mutation from the *R-sc:86* line to the *R-sc:86-17pale* subline. However, the behavior of the kernels in successive generations after the initial mutation is abnormal when considered from the perspective of a mutation causing each of the kernel color changes. The hypothesis that progressive kernel color changes result from a series of sequence mutations would require an improbable number and frequency of consecutive mutations to cause the stepwise changes in kernel color shown by K. McWhirtter and W. Eggleston (pers. comm.) (Figure 8). The speed of kernel color change, the consistency of the changes, the reversibility of the kernel color changes with selection, and the lack of sequence differences in the region found to be critical to the kernel color change all argue against the idea that the molecular basis of the kernel color changes are based on an ongoing series of sequential mutations (Meins and Lutz 1979). Another possible explanation that is related to source of *R-sc:86* is the idea that the kernel color changes in the *R-sc:86-17pale* sublines may be caused by transposable elements (TEs) in and around the *Sc||nc1* gene (Weil and Martienssen 2008). As described in Chapter one, TEs were an integral part of the expression pattern of *R-stippled*, and are present in the *Nc2* and *Nc3* genes of the *R-sc:86* line, knocking out expression of the *Nc2* and *Nc3* genes (Table 1) (Figure 7) (W. Eggleston, pers. com.) (Alleman and Kermicle

1993). However, the hypothesis of TEs being the underlying molecular basis for the kernel color changes in the *R-sc:86-17pale* sublines has several problems; the differences in expression of the kernel color would require a very high level of insertions, deletions, or changes in the TEs themselves. The recombination mapping performed by W. Eggleston (pers. comm.) shows that recombination in the 3` end of the second intron of the *Sc||nc1* gene occurred in all of the all 11 revertants to full color tested, identifying the 3` end of the second intron of the *Sc||nc1* gene being necessary for the continuum of the kernel color changes (Derkits 2013). With the sequencing results of the 3` end of the intron of the *Sc||nc1* gene showing that *R-sc:86*, *R-sc:86-17pale* Lightest, and *R-sc:86-17pale* Darkest having identical sequences in the area found to be of interest, TEs causing the kernel color change are not a viable explanation of the molecular basis for the kernel color changes. Likewise the possible idea of a distant mutation or series of mutations being the molecular basis for phenotypic changes, as was found to be the case for the *B* locus (Arteaga-Vazquez & Chandler 2010). The recombination mapping that found only the lightest kernel color or the full color revertants that had a recombination event at the 3` end of the *Sc||nc1* gene argues against both a series of affectors of kernel color, and against a single distant affector of kernel color (Stinard, Kermicle, and Sach 2008). A related hypothesis for the molecular basis of the kernel color change is that RNAi is affecting the expression levels of the *Sc||nc1* gene (McGinnis *et al.* 2007). This hypothesis cannot currently be proven or disproven for several reasons, as the generational testing for this hypothesis has some problems due to the non-viability of maize with the *mop1-* and *rmr6-* past three generations. W. Eggleston (pers. comm.) performed several sets of crosses that knocked out or down the expression of genes

critical for the expression and function of RNAi in maize, and found that *R-sc:86-17pale* Lightest in combination with the *mop1-* and *rmr6-* crosses did not show a significant change in kernel color, but did show slight increases in kernel mottling in the second and third generations (Figure 12). While it is possible that the RNAi effects may not have been removed in the three generations after inclusion with the *mop1-* and *rmr6-* mutations, the lack of kernel color change when RNAi regulatory mechanisms were removed is strong evidence that the molecular basis of the kernel color changes is not entirely RNAi based (Woodhouse 2006). Related to the use of *mop1-* and *rmr6-* is the hypothesis that paramutation is a possible molecular basis for the kernel color change (Erhard & Hollick 2011).

Also epigenetic in nature is the hypothesis that the molecular basis of the kernel color change is habituation, or heritable changes in expression in response to environmental pressure that are not based in sequence changes. As Meins and Lutz (1979) showed with tissue cultured tobacco cells, habituation does not rely on genetic change to cause phenotypic changes. This is supported by the work of Leroy *et al.* (2017), who showed in *Arabidopsis* that methylation may be connected with canalization states and changes in phenotypic expression. The sequencing results are evidence that there is not a sequence difference between the *R-sc:86* line, the *R-sc:86-17pale* Lightest, and the *R-sc:86-17pale* Darkest sublines, and the restriction digestion mapping and bisulfite sequencing suggest that the molecular basis of the kernel color is epigenetic in nature. The work presented here is not conclusive, but the sequencing and DNA cytosine methylation data are evidence in support of epigenetic differences, and that the epigenetic differences are associated with the kernel color changes (Figure 26).

Future Directions and Relevance

The results of this project are not conclusive, but the results do suggest that the molecular basis for the kernel color changes is epigenetic in nature, supporting earlier work by W. Eggleston (pers. comm.). Further work on this system could yield more information and conclusive results, but will need more work for the methylation data to become clearer, and will need to use newer methods of inquiry to address such epigenetic mechanisms as histone and chromatin mehtylation than were available at the beginning of the project (as reviewed in Bewick *et al.* 2017). While more work using the current sodium bisulfite treatment, PCR, and sequencing methods do have the potential to contribute further, the methods do not yield data consistently, and have problems with reproducibility in terms of bisulfite treatment success. The implications of this are that the current methods can be used to collect more data, but the samples used will require a significant investment both of time and of further materials cost in order to collect further data (Henderson, *et al.* 2010). This inconsistency may be reduced by the use of DNA samples extracted (peeled) solely from aluerone tissue, which is difficult to isolate, but may provide a sample without any trace starch that may be interfering with the bisulfite conversion. However, two alternative methodologies exist that hold promise for future study of the methylation patterns of the *R-sc:86* line, the *R-sc:86-17pale* Lightest subline, and the *R-sc:86-17pale* Darkest subline. One is Whole Genome Bisulfite Sequencing (WGBS), which has been used successfully for creating methylomes of maize, and may be useful for methylome analysis, but has the downside of requiring large amounts of sequencing to get incomplete coverage of the genome, and of using a selection of

methylation sites, rather than examining all of the methylation sites (Springer 2018). Another option is the use of the MinIon sequencing system, which would allow the sequencing of large portions of the genome, but requires DNA samples that have a higher purity than are currently available, and is not currently as accurate as Sanger sequencing, requiring a large number of sequencing reads for accuracy (Tyler 2018). As the cost of sequencing continues to decrease, the use of such techniques as the MinIon sequencing system will become more viable, and will make the continuation of this project possible. Another option that was not available when the project was started are more accurate methods that can be used to examine the chromatin and histones in the region of interest. These increasing functionality of these new techniques make not just further research into this project more viable, they also make it possible to further apply the study of epigenetics to crop research, and to reduce the developmental time frame for development of new plant varieties (Kuchel *et al.* 2005, Yeu 2014, Teh *et al.* 2017).

Literature Cited:

Alleman, M. & J. Kermicle. 1993. Somatic Variegation and Germinal Mutability Reflect the Position of Transposable Element *Dissasociation* Within the Maize *R* gene. *Genetics* 135:189-203.

Arteaga-Vazquez M.A. and V.L. Chandler. 2010. Paramutation in maize: RNA mediated trans-generational gene silencing. *Curr Opin Genet Dev* 20:156-163.

Ashman R.B. 1960. Stippled aleurone in maize. *Genetics* 45:19-19-34.

Averitt B., C. Shang, L. Rosso, J. Qin, M. Zhang, K.M. Rainy, and B. Zhang. 2017. Impact of *mips1*, *lpa1*, and *lpa2* Alleles for Low Phytic Acid Content on Agronomic, Seed Quality, and Seed Composition Traits of Soybean. *Crop Sci* 57:2490-2499.

Bako A., G. Gell, and E. Balazs. 2011. Quantification of transgene expression in maize (Zea mays L.) throughout the vegetation period. *Plant Breed* 130:41-45.

Banerjee R., J. Hasler, R. Meagher, R. Nagoshi, L. Hietala, F. Huang, K. Narva, and J.L. Jurat-Fuentes. 2017. Mechanism and DNA-based detection of field-evolved resistance to transgenic Bt corn in fall armyworm (Spodoptera frugiperda). *Scientific Reports* 7:10877.

Bedford T., A. Rambaut, and M. Pascual. 2012. Canalization of the evolutionary trajectory of the human influenza virus. *Bmc Biology* 10:38.

Bewick A.J., C.E. Niederhuth, L. Ji, N.A. Rohr, P.T. Griffin, J. Leebens-Mack, and R.J. Schmitz. 2017. The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. *Genome Biol* 18:65.

Bewick, A. and R. Schmitz. 2017. Gene body DNA methylation in plants. *Current Opinion in Plant Biology* 36:v103-110.

Bustos F.J., N. Jury, P. Martinez, E. Ampuero, M. Campos, S. Abarzua, K. Jaramillo, S. Ibing, *et al.* 2017. NMDA receptor subunit composition controls dendritogenesis of hippocampal neurons through CAMKII, CREB-P, and H3K27ac. *J Cell Physiol* 232:3677-3692.

Bygren L.O., P. Tinghog, J. Carstensen, S. Edvinsson, G. Kaati, M.E. Pembrey, and M. Sjostrom. 2014. Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren. *Bmc Genetics* 15:12.

Carey C.E., A. Agrawal, B. Zhang, E.D. Conley, L. Degenhardt, A.C. Heath, D. Li, M.T. Lynskey, *et al.* 2015. Monoacylglycerol Lipase (MGLL) Polymorphism rs604300 Interacts With Childhood Adversity to Predict Cannabis Dependence Symptoms and Amygdala Habituation: Evidence From an Endocannabinoid System-Level Analysis. *J Abnorm Psychol* 124:860-877.

Chanderbali A. S., M. Yoo, L. M. Zahn, S. F. Brockington, P. K. Wall *et al.* 2010. Conservation and canalization of gene expression during angiosperm diversification accompany the origin and evolution of the flower. *Proc. Natl. Acad. Sci. U. S. A.* 107:22570-22575.

Chen J., I. Greenblatt, and S. Dellaporta. 1992. Molecular Analysis of Ac Transposition and Dna-Replication. *Genetics* 130:665-676.

Chen Y., T. Liu, C. Yu, T. Chiang, and C. Hwang. 2013. Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. *Plos One* 8:e62856.

Cifuentes-Zuniga F., V. Arroyo-Jousse, G. Soto-Carrasco, P. Casanello, R. Uauy, B.J. Krause, and J.A. Castro-Rodriguez. 2017. *IL-10* expression in macrophages from

neonates born from obese mothers is suppressed by *IL-4* and *LPS/INF*. *J Cell Physiol* 232:3693-3701.

Clark R., S. Tavare, and J. Doebley. 2005. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Mol Biol Evol* 22:2304-2312.

Day T. 2016. Interpreting phenotypic antibiotic tolerance and persister cells as evolution via epigenetic inheritance. *Mol Ecol* 25:1869-1882.

de Castro M., R. Martinez-Rubio, J.L. Acebes, A. Encina, S.C. Fry, and P. Garcia-Angulo. 2017. Phenolic metabolism and molecular mass distribution of polysaccharides in cellulose-deficient maize cells(FA). *J Integr Plant Biol* 59:475-495.

Dietrich, W. 1993. Determination of rates and patterns of recombination at the maize *red color* (*r1*) locus. Master's Thesis. Virginia Commonwealth University.

Derkits J. 2013. Genetic and molecular analysis of canalization at the maize *r1* locus. Master's Thesis. Virginia Commonwealth University.

Dooner, H. K., T. P. Robbins, and R. A. Jorgense. 1992. Genetic and developmental control of anthocyanin biosynthesis. *Annu. Rev. Genet.* 25:173-200.

dos Santos C., V. Dalpasquale, C. Scapim, A. Braccini, and F. Krzyzanowski. 2005. Milk line as an indicator of the harvesting time of three hybrid seeds of corn (Zea mays L.). *Braz Arch Biol Technol* 48:161-170.

Eggleston, W. B., M. Alleman and J. L. Kermicle. 1995. Molecular organization and germinal instability of R-stippled maize. *Genetics* 141:347-360.

Eichten S.R., R.A. Swanson-Wagner, J.C. Schnable, A.J. Waters, P.J. Hermanson, S. Liu, C. Yeh, Y. Jia, *et al.* 2011. Heritable Epigenetic Variation among Maize Inbreds. *PLoS Genet* 7:e1002372.

Eichten, S., R. Brickine, J. Song, Q. Li, R. Swanson-Wagner, P. Hermanson, A. Waters, E Starr, *et al.* 2013. Epigenetic and genetic influences on DNA methylation variation in maize populations. *Plant Cell* 25: 2783–2797.

Erhard, K., and J. Hollick. 2011. Paramutation: a process for acquiring *trans*-generational regulatory states. *Current Opinion in Plant Biology* 14:210-216.

Etchevers G., O. Banasik, and C. Watson. 1976 A Method for Instumental Measurement of Barley Color. *Cereal Chemistry* 53(6):846-853.

Eyster, W. 1925. The Effect of Evironment on Variegation Patterns in Maize Pericarp. *Genetics* 10(2):179-196.

Fanti L., L. Piacentini, U. Cappucci, A.M. Casale, and S. Pimpinelli. 2017. Canalization by Selection of de Novo Induced Mutations. *Genetics* 206:1995-2006.

Fisher J., E. Bensal, and D. Zamir. 2017. Bimodality of stable and plastic traits in plants. *Theor Appl Genet* 130:1915-1926.

Gardner A., C. Tchou, T Hawkins, P Debevec. 2003. Linear Light Source Reflectometry. *ACM Transactions on Graphics* 22:749-758.

Haberer G., S. Young, A. Bharti, H. Gundlach, C. Raymond, G. Fuks, E. Butler, R. Wing, *et al.* 2005. Structure and architecture of the maize genome. *Plant Physiol* 139:1612-1624.

Han, Z., P. Crisp, S. Stelpflug, S. Kaeppler, Q. Li, and N. Springer. 2018. Heritable Epigenomic Changes to the Maize Methylome Resulting from Tissue Culture. *Genetics* 209:983-995.

Helentjaris T., M. Slocum, S. Wright, A. Schaefer, and J. Nienhuis. 1986. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymophisms. *Theor. Appl. Genet.* 72(6):761-769.

Henderson, I., S. Chan, X. Cao, L. Johnson, and S. Jacobsen. 2010. Accurate sodium bisulfite sequencing in plants. *Epigenetics* 5(1):47-49.

Hollick, J., J. Kermicle, and S. Parkinson. 2005. *Rmr6* Maintains Meiotic Inheritance of Paramutant States in *Zea mays*. *Genetics* 171:725-740.

Hornstein E. and N. Shomron. 2006. Canalization of development by microRNAs. *Nat Genet* 38:S20-S24.

Hsu, F., M. Yen, C. Wang, C. Lin, C. Wang, and P. Chen. 2017. Optimized reduced representation bisulfite sequencing reveals tissue-specific mCHH islands in maize. *Epigenetics Chromatin* 10:42

Hsu F., C. Wang, and P. Chen. 2013. Reduced Representation Bisulfite Sequencing in Maize. *Bio-Protocol* 8(6):1

Kieleczawa J. 2006. Fundamentals of sequencing of difficult templates--an overview. *Journal of biomolecular techniques : JBT* 17:207-17.

Kishi R., A. Araki, M. Minatoya, T. Hanaoka, C. Miyashita, S. Itoh, S. Kobayashi, Y.A. Bamai, *et al.* 2017. The Hokkaido Birth Cohort Study on Environment and Children's Health: cohort profile-updated 2017. *Environmental Health and Preventive Medicine* 22:46.

Kuchel H., G. Ye, R. Fox, and S. Jefferies. 2005. Genetic and economic analysis of a targeted marker-assisted wheat breeding strategy. *Mol Breed* 16:67-78.

Kusmec A., S. Srinivasan, D. Nettleton, and P.S. Schnable. 2017. Distinct genetic architectures for phenotype means and plasticities in *Zea mays*. *Nature Plants* 3:715-723.

Lamarck J.B. 1809. Philosophie Zoologique. Museum d'Histoire Naturelle, Paris.

Largo-Gosens A., A. Encina, M. de Castro, H. Melida, J.L. Acebes, P. Garcia-Angulo, and J.M. Alvarez. 2016. Early habituation of maize (*Zea mays*) suspension-cultured cells to 2,6-dichlorobenzonitrile is associated with the enhancement of antioxidant status. *Physiol Plantarum* 157:193-204.

Le Roy A., I. Loughland, and F. Seebacher. 2017. Differential effects of developmental thermal plasticity across three generations of guppies (Poecilia reticulata): canalization and anticipatory matching. *Scientific Reports* 7:4313.

Li H., M. Freeling, and D. Lisch. 2010. Epigenetic reprogramming during vegetative phase change in maize. *Proc Natl Acad Sci U S A* 107:22184-22189.

Li Q., J.I. Gent, G. Zynda, J. Song, I. Makarevitch, C.D. Hirsch, C.N. Hirsch, R.K. Dawe, *et al.* 2015. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proc Natl Acad Sci U S A* 112:14728-14733.

Lindsay, R. and W. Eggleston. 2014. Improved DNA extraction from high starch maize tissue using a Sodium Dodecyl Sulfate extraction method. *Maize Journal of India* 3:18-23

Ludwig, S. R., B. Bowen, L. Beach and S. R. Wessler. 1990. A regulatory gene as a novel visible marker for maize transformation. *Science* 247:449-450.

Ludwig, S., L. Habera, S. Dellaporta and S. Wessler. 1989. *Lc,* a member of the maize R-gene family responsible for tissue-specific anthocyanin production, encodes a protein

similar to transcriptional activators and contains the myc-homology region. *Proc. Natl. Acad. Sci. U. S. A.* 86:7092-7096.

Mach, J. 2013. Epigenetics and Genetics, Global profiling in Maize inbred lines examines variation in DNA methylation. *Plant Cell* 25(8): 2769.

Maroof M.A.S., N.M. Glover, R.M. Biyashev, G.R. Buss, and E.A. Grabau. 2009. Genetic Basis of the Low-Phytate Trait in the Soybean Line CX1834. *Crop Sci* 49:69-76.

Macnair M.R. 1987. Heavy-Metal Tolerance in Plants - a Model Evolutionary System. *Trends in Ecology & Evolution* 2:354-359.

Martinez J. and F. Baquero. 2000. Mutation frequencies and antibiotic resistance. *Antimicrob Agents Chemother* 44:1771-1777.

Marquez B., G. Ameye, C.M. Vallet, P.M. Tulkens, H.A. Poirel, and F. Van Bambeke. 2011. Characterization of Abcc4 Gene Amplification in Stepwise-Selected Mouse J774 Macrophages Resistant to the Topoisomerase II Inhibitor Ciprofloxacin. *Plos One* 6:e28368.

McGinnis, K., N. Murphy, A. R. Carlson, A. Akula, C. Akula, H. Basinger, M. Carlson, *et al.* Assessing the Efficiency of RNA Interference for Maize Functional Genomics. *Plant Physiology* 143:1441-1451.

McNamara J.M., S.R.X. Dall, P. Hammerstein, and O. Leimar. 2016. Detection vs. selection: integration of genetic, epigenetic and environmental cues in fluctuating environments. *Ecol Lett* 19:1267-1276.

McWhirter K.S. 1961. Paramutation studies with self-colored mutants from the allele in maize.

Meins F. 1989. Habituation - Heritable Variation in the Requirement of Cultured Plant-Cells for Hormones. *Annu Rev Genet* 23:395-408.

Meins F. and J. Lutz. 1979. Tissue-specific variation in the Cytokinen Habituation of Cultured Tobacco Cells. *Differentiation* 15:1-6.

Mendizabal I., T.E. Keller, J. Zeng, and S.V. Yi. 2014. Epigenetics and Evolution. *Integrative and Comparative Biology* 54:31-42.

Milton C.C., C.M. Ulane, and S. Rutherford. 2006. Control of Canalization and Evolvability by Hsp90. *Plos One* 1:e75.

Mischo H.E. and N.J. Proudfoot. 2013. Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* 1829:174-185.

Monjardino, P., A. Smith, and R. Jones. 2006. Zein Transcription and Endoreduplication in Maize Endosperm are Differentially Affected by Heat Stress. *Crop Science* 46:2581-2589.

Okagai R., S. Dukowic-Schulze, W. Eggleston, and G. Muehlbauer. 2018. A Critical Assessment of 60 Years of Maize Intragenic Recombination. *Front. Plant Sci.* 9:1560

Okonechinikov k., O. Golosova, M. Fursov, and The UGENE team. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28:1166.

Petroni, K., E. Cominelli, G. Consonni, G. Gusmaroli, G. Gavazzi, C. Tonelli. 2000. The Developmental Expression of the Maize Regulatory Gene *Hopi* Determines Germination Dependent Anthrocyanin Accumulation. *Genetics* 155(1):323-336.

Phillips, R. L. 2010. Mobilizing science to break yield barriers. *Crop Sci.* 50:S99-S108.

Ponomarev I., S. Wang, L. Zhang, R.A. Harris, and R.D. Mayfield. 2012. Gene Coexpression Networks in Human Brain Identify Epigenetic Modifications in Alcohol Dependence. *J Neurosci* 32:1884-1897.

Queitsch, C., T. Sangster and S. Lindquist. 2002. *Hsp90* as a capacitor of phenotypic variation. *Nature* 417:618-624.

Ross M.G., C. Russ, M. Costello, A. Hollinger, N.J. Lennon, R. Hegarty, C. Nusbaum, and D.B. Jaffe. 2013. Characterizing and measuring bias in sequence data. *Genome Biol* 14:R51.

Rutherford, S., and S. Lindquist. 1998. *Hsp90* as a capacitor for morphological evolution. **Nature** 396:336-342.

Salathia N. and C. Queitsch. 2007. Molecular mechanisms of canalization: *Hsp90* and beyond. *J Biosci* 32:457-463.

Salehi M., E. Kamali, M. Karahmadi, and S.M. Mousavi. 2017. RORA and Autism in Isfahan Population: Is There An Epigenetic Relationship. *Cell Journal* 18:540-546.

Sato, K., and H. Siomi. 2010. Is canalization more than just a beautiful idea? *Genome Biol*. 11:109.

Savickiene J., D. Matuzevicius, S. Baronaite, G. Treigyte, N. Krasovskaja, I. Zaikova, D. Navakauskas, A. Utkus, and R. Navakauskiene. 2017. Histone Modifications Pattern Associated With a State of Mesenchymal Stem Cell Cultures Derived From Amniotic Fluid of Normal and Fetus-Affected Gestations. *J Cell Biochem* 118:3744-3755.

Sieriebriennikov B., G.V. Markov, H. Witte, and R.J. Sommer. 2017. The Role of *DAF-21/Hsp90* in Mouth-Form Plasticity in *Pristionchus pacificus*. *Mol Biol Evol* 34:1644-1653.

Springer, N., S Anderson, C Andorf, K. Ahern, F. Bai, O. Barad, W. Barbazuk, H. Bass, *et al.* 2018. The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nature Genetics* 50:1282-1291.

Stadler L. 1946. Spontaneous Mutation at the R-Locus in Maize. I. the Aleurone-Color and Plant-Color Effects. *Genetics* 31:377-394.

Stam M., C. Belele, J. Dorweiler, and V. Chandler. 2002. Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. *Genes Dev* 16:1906-1918.

Stinard, P., J. Kermicle, and M. Sachs. 2008. The Maize *enr* System of *r1* Haplotype-Specific Aleurone Color Enhancement Factors. *Journal of Heredity* 100(2):217-228.

Tang B., R. Zhao, Y. Sun, Y. Zhu, J. Zhong, G. Zhao, and N. Zhu. 2011. Interleukin-6 expression was regulated by epigenetic mechanisms in response to influenza virus infection or dsRNA treatment. *Mol Immunol* 48:1001-1008.

Tukey J. 1949. Comparing Individual Means in the Analysis of Variance. *Biometrics* 5 (2):99-114.

Turco G., K. Kajala, G. Kunde-Ramamoorthy, C. Ngan, A. Olsen, S. Deshphande, D. Tolkunov, *et al.* 2017. DNA methylation and gene expression associated with vascularization in *Sorgum bicolor*. *New Phytol.* 214(3)1213-1229.

Tyler A., L Mataseje, C. Urfano, L. Schmidt, K. Antonation, M. Mulvey, and C. Corbett. 2018. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Scientific Reports* 8:10931.

Vergeer, P., N. Wagemaker, and N. Ouborg. 2012. Evidence for an Epigenetic role in inbreeding depression. *Biol Letters* 8(5):798-801.

Waddington C. 1942. Canalization of Development and the Inheritance of acquired Characters. *Nature* 150:563.

Waddington, C. 1956 *Principles of Embryology*. Macmillian, New York.

Walker, E. 1998. Paramutation of the *r1* Locus of Maize is Associated with Increased Cytosine Methlyation. *Genetics* 148:1973-1981.

Wang L., H. Liu, D. Li, and H. Chen. 2011. Identification and characterization of maize microRNAs involved in the very early stage of seed germination. *BMC Genomics* 12:154.

Wang X., A.A. Elling, X. Li, N. Li, Z. Peng, G. He, H. Sun, Y. Qi, X.S. Liu, and X.W. Deng. 2009. Genome-Wide and Organ-Specific Landscapes of Epigenetic Modifications and Their Relationships to mRNA and Small RNA Transcriptomes in Maize. *Plant Cell* 21:1053-1069.

Weil, C., & R. Martienssen. 2008. Epigenetic interactions between transposons and genes: lessons from plants. *Current opinions in Genetics and Development* 18:188-192.

Werner R.J., A.D. Kelly, and J.J. Issa. 2017. Epigenetics and Precision Oncology. *Cancer Journal* 23:262-269.

West P.T., Q. Li, L. Ji, S.R. Eichten, J. Song, M.W. Vaughn, R.J. Schmitz, and N.M. Springer. 2014. Genomic Distribution of H3K9me2 and DNA Methylation in a Maize Genome. *PLoS One* 9:e105267.

Woodhouse M., M. Freeling, and D. Lisch. 2006. The *mop1 (mediator of paramutation)* mutant progressively reactivates one of the two genes encoded by the *MuDR* Transposon in Maize. *Genetics* 172(1):579-592.

Yue G.H. 2014. Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish Fish* 15:376-396.

Appendix 1, Tables

Allele	Parent	Haplotypes	3' end Methylation of *r1* genes (CG)	3' end Methylation of *r1* genes (CNG)	Kernel Color	Source
r-g:Δ902	*R-sc:N1-575-1/r-r:n46*	No *r1* groups	N/A	N/A	colorless	(Kermicle 1985; Alleman and Kermicle 1993)
R-stippled	Land race	*Sc Nc1 Nc2 Nc3*	-	-	(spotted)	(Eggleston *et al.* 1995)
R-sc:86	*R-Stippled*	*Sc||nc1 Nc2 Nc3*	-	-	black/ purple	(Ashman 1960; Eggleston *et al.* 1995)
R-sc:86-17pale	*R-sc:86*	*Sc||nc1 Nc2 Nc3*	++	+	medium	(W. Eggleston pers. comm.)
R-sc:86-17pale LLLLL (lightest	*R-sc:86-17pale*	*Sc||nc1 Nc2 Nc3*	++++	++++	colorless aleurone	(W. Eggleston pers. comm.)
R-sc:86-17pale DDDDD darkest	*R-sc:86-17pale*	*Sc||nc1 Nc2 Nc3*	+	++	black/ purple	(W. Eggleston pers. comm.)
R-sc:124	*R-Stippled / r-g*	*Sc||nc3*	-	-	black/ purple	(Ashman 1960; Eggleston *et al.* 1995)

Table 1. **Origin, structure, and methylation of select *r1* haplotypes**. All alleles are in maize W22 inbred line. Methylation status determined by Southern blot analysis of methylation patterns (W. Eggleston pers. comm.) (Figure 10).

Line	Planting delay	Alleles and Epitypes	*n*	Shapiro-Wilk	p-value
2011-076	0	*R-sc:86*	17	0.91	0.1021
2011-077	21 days	*R-sc:86*	26	0.97	0.6807
2011-078	0 days	*R-sc:86-17:* pale Darkest	21	0.94	0.1804
2011-079	21 days	*R-sc:86-17:* pale Darkest	26	0.95	0.1795
2011-080	0 days	*R-sc:86-17:* pale Lightest	10	0.83	0.03741*
2011-081	21 days	*R-sc:86-17:* pale Lightest	4	0.75	0.04086*
2012-104	0 days	*R-sc:86-17:* pale Med-Light	8	0.92	0.3979
2012-105	0 days	*R-sc:86-17:* pale Med-Dark	22	0.99	0.985
2012-106	0 days	*R-sc:86-17:* pale Darkest	2	na	na
2012-109	14 days	*R-sc:86-17:* pale Lightest	1	na	na
2012-110	14 days	*R-sc:86-17:* pale Med-Light	4	0.85	0.2399
2012-111	14 days	*R-sc:86-17:* pale Med-Light	4	0.89	0.3891
2012-112	14 days	*R-sc:86-17:* pale Med-Light	2	na	na
2014-327	0 days	*R-sc:86-17:* pale Darkest	9	0.86	0.1018
2014-328	0 days	*R-sc:86-17:* pale Med-Dark	11	0.85	0.03945*
2014-329	0 days	*R-sc:86-17:* pale Med-Light	4	0.95	0.7332
2014-330	0 days	*R-sc:86-17:* pale lightest	9	0.96	0.804
2014-331	0 days	*R-sc:124*	19	0.98	0.8881
2014-332	0 days	*r-g:*Δ902	17	0.93	0.245

Table 2. **Shapiro-Wilk test of normality of each of the lines.** Shapiro-Wilk and *n (#* of ears) values are shown for interpretation of the p-value given. p-values of less than 0.05 indicate a divergence from normal distribution in the line tested (*).

Comparison	Difference in means	Lower	Upper	Adjusted p-value
Darkest vs. *R-sc:86*	2.00	1.41	2.60	<<0.001
Lightest vs. *R-sc:86*	37.07	36.15	37.98	<<0.001
Lightest vs. Darkest	35.06	34.16	35.97	<<0.001

Table 3. **Tukey's Honest-significant differences test of light reflectance for lines grown in 2011.** The post hoc Tukey's HSD test shows the significant differences between all lines sampled in 2011, which have an adjusted p-value of less than the significance levels of 0.05. Difference in Means, Upper, and Lower are as described in Table 2.

Comparison	Difference in means	Lower	Upper	Adjusted p-value
Lightest vs. Darkest	28.55	10.04	47.06	<<0.01
Med-dark vs. Darkest	13.66	2.49	24.82	0.01
Med-light vs. Darkest	15.66	4.39	26.93	<0.01
Med-dark vs. Lightest	-14.89	-30.35	0.57	0.06
Med-light vs. Lightest	-12.89	-28.42	2.64	0.13
Med-light vs. Med-Dark	2.00	-2.804	6.81	0.68

Table 4. **Tukey's Honest Significant differences test of light reflectance for lines grown in 2012.** The post hoc Tukey's HSD test shows significant differences between Lightest and Darkest, Medium Dark and Darkest, and Medium Light and Darkest lines. The Tukey' HSD test did not show an honest significant difference between the Lightest, Medium Light, and Medium Dark lines. Difference in Means, Upper, and Lower are as described in Table 2.

Comparison	Difference in means	Lower	Upper	Adjusted p-value
*r-g:*Δ902 vs. Darkest	35.14	33.41	36.87	<<0.001
Lightest vs. Darkest	33.63	31.65	35.60	<<0.001
Med-Dark vs. Darkest	23.37	21.44	25.30	<<0.001
Med-Light vs. Darkest	24.82	22.30	27.34	<<0.001
R-sc:124 vs. Darkest	-3.42	-5.12	-1.72	<<0.001
Lightest vs. *r-g:*Δ902	-1.51	-3.24	0.22	0.120
Med-Dark vs. *r-g:*Δ902	-11.77	-13.44	-10.10	<<0.001
Med-Light vs. *r-g:*Δ902	-10.32	-12.65	-7.99	<<0.001
R-sc:124 vs. *r-g:*Δ902	-38.56	-39.96	-37.16	<<0.001
Med-Dark vs. Lightest	-10.26	-12.18	-8.33	<<0.001
Med-Light vs. Lightest	-8.81	-11.33	-6.29	<<0.001
R-sc:124 vs. Lightest	-37.05	-38.74	-35.35	<<0.001
Med-Light vs. Med-Dark	1.449	-1.03	3.93	0.525
R-sc:124 vs. Med-Dark	-26.79	-28.43	-25.15	<<0.001
R-sc:124 vs. Med-Light	-28.24	-30.55	-25.93	<<0.001

Table 5. **Tukey's Honest Significant differences test of light reflectance of kernels from 2014.** The post hoc Tukey's HSD test shows significant differences between all samples tested, except between the Lightest to *r-g:Δ902*, and the Medium Dark to Medium Light comparisons at the 0.05 significance level. Difference in Means, Upper, and Lower are as described in Table 2.

Sequence bp	Subline with sequence difference	Consensus sequence	Sequence difference	Location
284	Darkest	A	N	Intron 2
618	Lightest	G	N	Exon 3
1227	Darkest	C	N	Exon 5
1283	Darkest	C	T	Intron 5
1393	Darkest	C	N	Intron 6
1410	Darkest	C	G	Intron 6

Table 6. **Potential sequence differences between the *R-sc:86-17pale* Lightest, *R-sc:86-17pale* Darkest, and *R-sc:86* lines.** Six potential sequence differences from the consensus sequence were found in the Lightest, Darkest, and *R-sc:86* consensus sequences (Figure 18, Figure 19, Figure 20).

Primer	Sequence location	Sequence	Nest
Sc3841a	3840-3857	5'-TTC TTC TCT ACC CTT CGC	-
Lc1933	4265-4250	5'-AGC AAG CTG GCT CCT C	-
Sc7737	7741-7759	5'-CTC GGG AAG GAG CGA GAA G	-
Mar140	8307-8285	5'-CTT CAC CTC GCC GTT GTA GAA CC	-
Lc8701	10714-10691	5'-GCC TTC CAT GCC CG TCG ATG TCC A	-
Sc8066a	8069-8087	5'-GGC TGC CGT GTG GAG CCA G	-
Sc8066b	8087-8069	5'- GCT CCA CAC GGC AGC C	-
Lc6713	8711-8727	5'-TCG CCG GCA GCA AAG CC	-
Lc6744	8759-8740	5'-GTG CAT ACC TTG GCC AGG AG	-
Lc7076	9074-9092	5'-TGT TCA GCT CGA GCT TCC G	-
Lc7062	9078-9060	5'-GCT GAA CAT ACC GTG TCA GTT G	-
Lc7295a	9310-9330	5'-GGC ACG TTT GCG TTC GAG GAA	-
Lc7497	9520-9503	5'-GCG GTA GTG GTA GCG CCT	-
Lc7775a	9806-9788	5'-CCA CAG CTC TCC CAA GCA C	-
Lc7775b	9788-9806	5'-GTG CTT GGG AGA GCT GTG G	-
Lc8087a	10100-10077	5'-GGG TGC AAG AGC TGG AGT CCA GT	-
Bs47527	7527-7546	5`-CTC CAA TTC CRA CRT CCA R	Outer 2
Bs47666m	7666-7684	5'-GTA AAA CGA CGG CCA GCT RCT AAA CRC TTR CTC CRC A	Outer 1, Inner 1,2
Bs2Sc8045r	8067-8048	5'CCR ACR TTA RTC ACR TTA CT	Inner 1, 2
Bs2Sc8144r	8219-8199	5`-CCT CAR RAA CAC TRR TTT C	Outer 2
Bs2Lc6237r	8261-8235	5'-CAA CRA CCA TAT TTT ATT TRA TRT CCT	Outer 1
m13F	-	5`-GTA AAA CGA CGG CCA G	-

Table 7. **Primers used for amplification of the region of the 3' end of the *Sc||nc1* gene.** Primer names, sequences, and annealing locations in the *Sc||nc1* gene, with nest details shown for all successful sodium bisulfite treated nest amplifications yielding the sequences shown in
Table 8 (Figure 9, Figure 21).

R-sc:124 Reference Numbering	Normalized guanidine site numbers	400-6e5 Darkest		504-11g1 Darkest		504-9g1 *R-sc:86*		496-3g1 Lightest		400-5e3 Lightest		Type of Methylation
		%G	%A	%G	%A	%G	%A	%G	%A	%G	%A	
7759	-15	100	0	100	0	0	100	100	0	100	0	CHH
7761	-13	100	0	100	0	100	0	100	0	100	0	CHH
7777	3	100	0	100	0	100	0	100	0	100	0	CHG
7778	4	100	0	100	0	100	0	100	0	100	0	CG
7781	7	100	0	100	0	100	0	100	0	100	0	CHH
7782	8	100	0	100	0	100	0	100	0	100	0	CHG
7787	13	95	5	100	0	100	0	40	60	100	0	CHH
7792	18	100	0	100	0	100	0	100	0	100	0	CHH
7794	20	90	10	100	0	100	0	100	0	100	0	CHH
7801	27	100	0	100	0	100	0	100	0	100	0	CHH
7802	28	100	0	100	0	100	0	100	0	100	0	CHH
7807	33	30	70	10	90	10	90	20	80	0	100	CHH
7813	39	65	35	60	40	55	45	65	35	50	50	CHH
7823	49	35	65	40	60	0	100	30	70	30	70	CHH
7825	51	70	30	100	0	100	0	100	0	80	20	CHH
7835	61	100	0	100	0	0	100	100	0	100	0	CHH
7836	62	100	0	100	0	100	0	100	0	100	0	CHH
7842	68	90	10	70	30	65	35	70	30	55	45	CHH
7847	73	95	5	100	0	100	0	100	0	100	0	CHH
7848	74	100	0	100	0	100	0	100	0	100	0	CHH
7875	101	100	0	100	0	100	0	100	0	100	0	CHH
7884	110	100	0	100	0	100	0	100	0	100	0	CHH
7885	111	100	0	100	0	100	0	100	0	100	0	CHH
7886	112	10	90	10	90	100	0	100	0	100	0	CHH
7888	114	100	0	100	0	100	0	100	0	100	0	CHG
7889	115	100	0	100	0	100	0	100	0	100	0	CG
7901	127	0	100	0	100	5	95	0	100	0	100	CG
7903	129	100	0	100	0	100	0	100	0	100	0	CHH
7904	130	100	0	100	0	100	0	100	0	100	0	CHH
7905	131	100	0	100	0	100	0	100	0	100	0	CHH
7906	132	100	0	100	0	100	0	100	0	100	0	CHH
7908	134	20	80	25	75	30	70	20	80	30	70	CHH
7909	135	90	10	100	0	95	5	95	5	100	0	CHH
7912	138	100	0	100	0	100	0	100	0	100	0	CHH

Table 8. **Tabulated methylation data for selected samples.** Base G # is set to a common reference point, which can be seen in Figure 22. %G and %A are the percentage of height of the peaks over the baseline for each of the bases in the electropherogram for the sample. The three types of methylation are color coded, with ChG methylation shown in green, CG methylation shown in yellow, and CHH methylation shown in light blue. Purple boxes mark sites where there is partial or no methylation at the site.

Cytosine methylation site type	**CG**	**CHH**	**CHG**	**Total**
total	3	25	4	32
Sites fully methylated	2	14	4	20
Sites not fully methylated	1	11	0	12
Unmethylated sites	0	0	0	0
% fully methylated sites	67%	56%	100%	
% partially methylated sites	33%	44%	0%	

Table 9. DNA cytosine methylation sites that are fully methylated vs. partially methylated on the bottom strand of the region sequenced.

Appendix 2, Figures

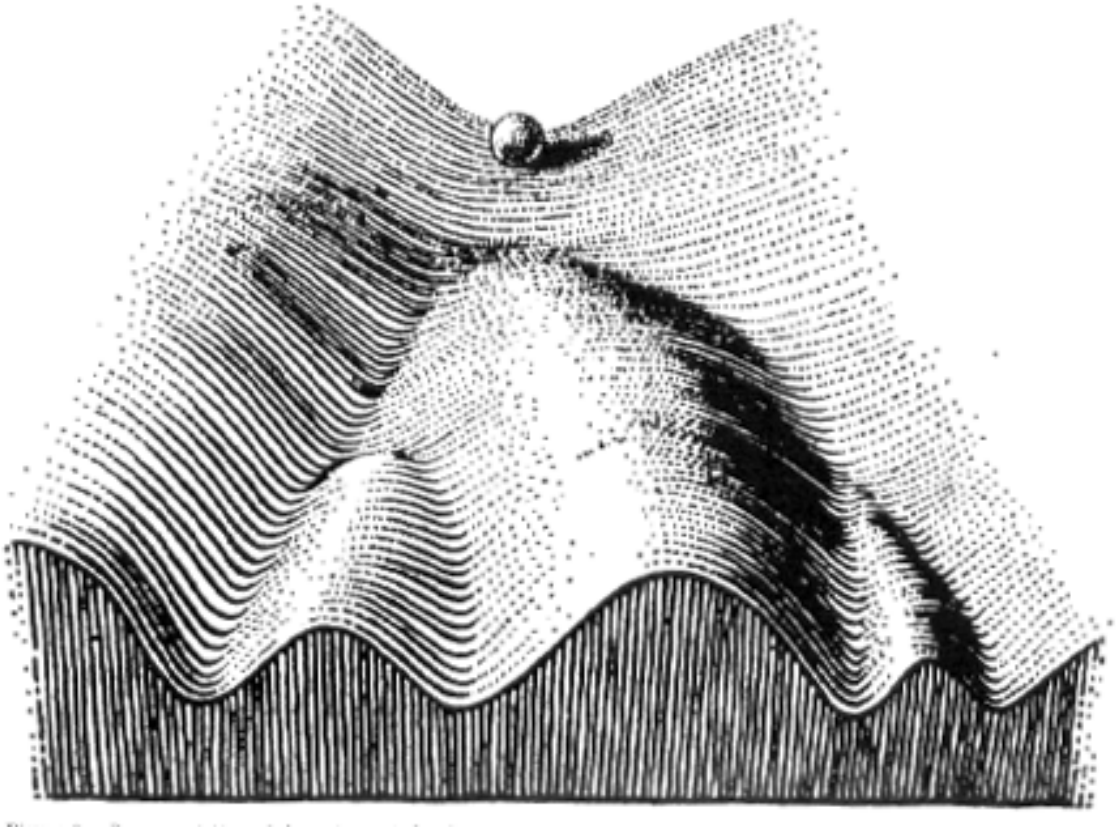


Figure 1. **Conceptual view of how canalization directs expression patterns.** From Waddington (1956).

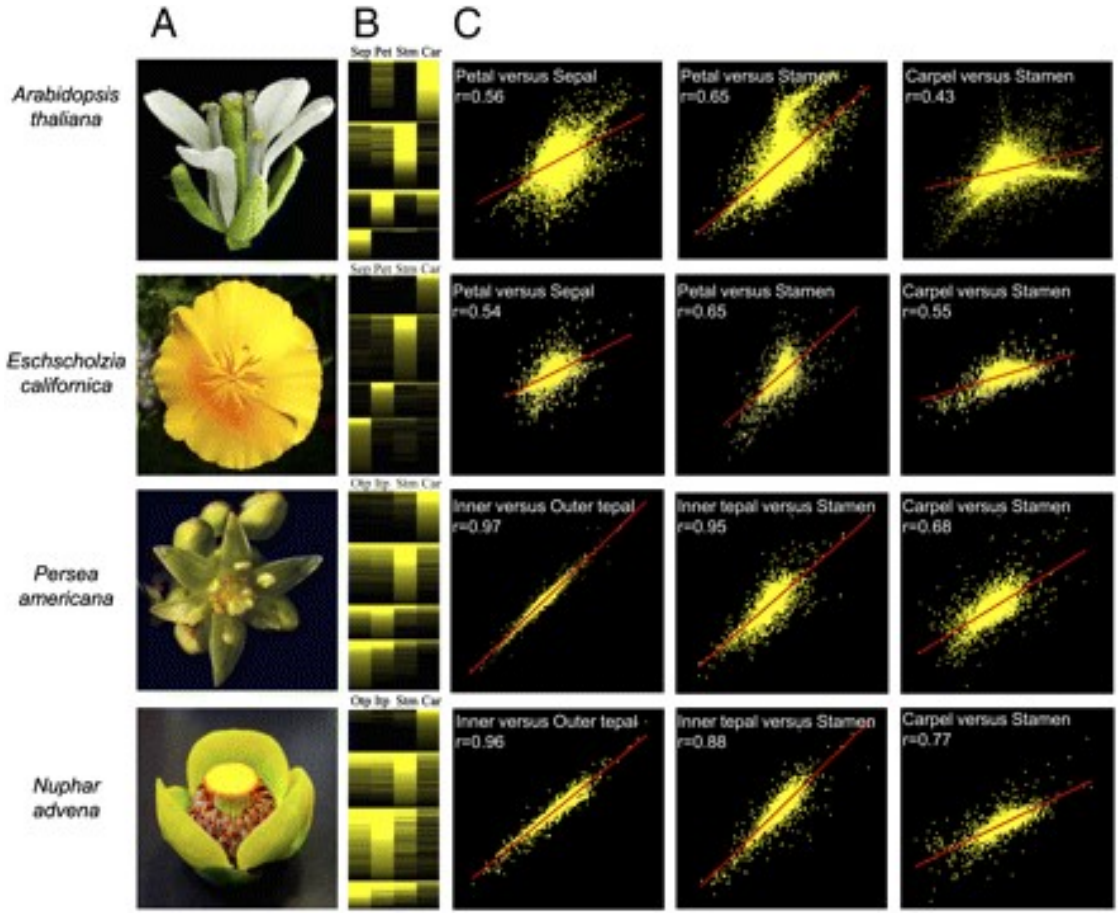


Figure 2. **A comparison of the possible range of expression vs. canalized expression patterns.** A) *Nuphar* and *Persea* contain an undifferentiated perianth of petaloid organs (tepals), whereas in *Eschscholzia* and *Arabidopsis* flowers the perianth is differentiated into leaf-like outer sepals and colorful inner petals. B) Log2 floral organ/leaf gene expression ratios ranked by organs of peak expression. C) Scatter plots of log2 floral organ/leaf ratios and Pearson correlations. From Chanderbali *et al.* (2010).

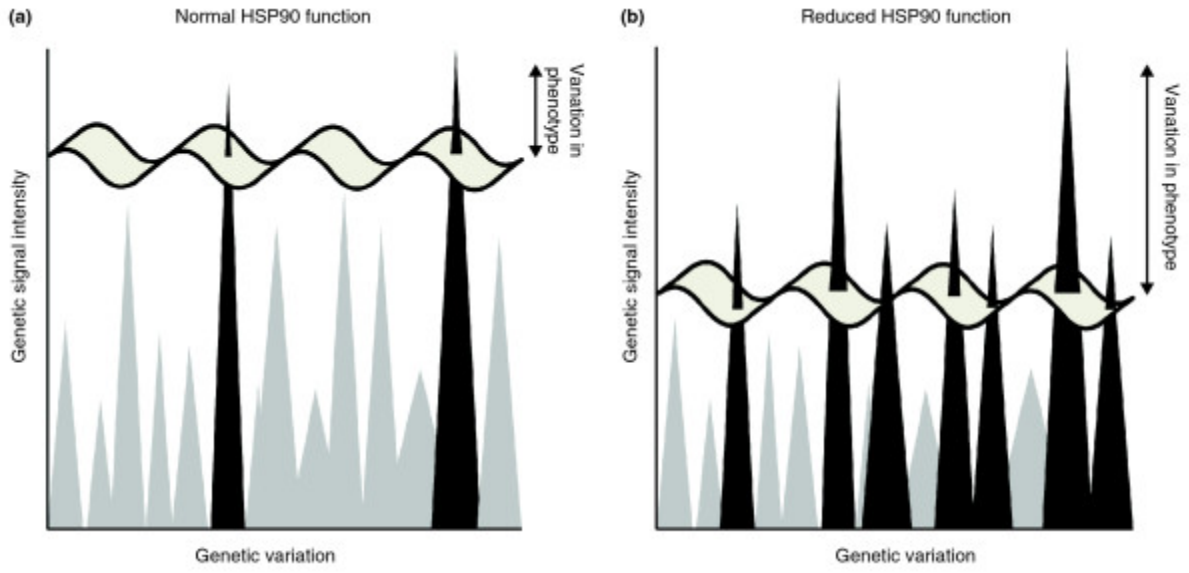


Figure 3. **A model for the buffering role of *Hsp90* in canalization**. **(a)** When Hsp90 is functional, underlying cryptic genetic variation (gray peaks) is hidden, and normally expressed genes (black peaks) are output as visible phenotypes. **(b)** When *Hsp90* is non-functional, previously cryptic phenotypic variation (formerly gray peaks) is expressed, as the phenotypic expression threshold (wavy line) is lowered. From Sato and Siomi (2010).

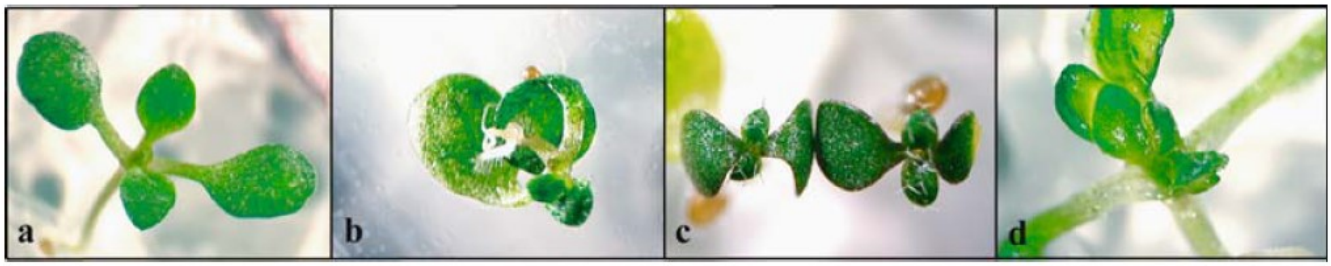


Figure 4. **A Comparison of *Arabadopsis thaliana* seedlings with *Hsp90* inhibition**. Panel A shows a healthy wild type 14 day post germination *A. thaliana* seedling, while panels B, C, and D show mutant phenotypes resulting from *Hsp90* inhibition. From Queitsch *et al.* (2002).

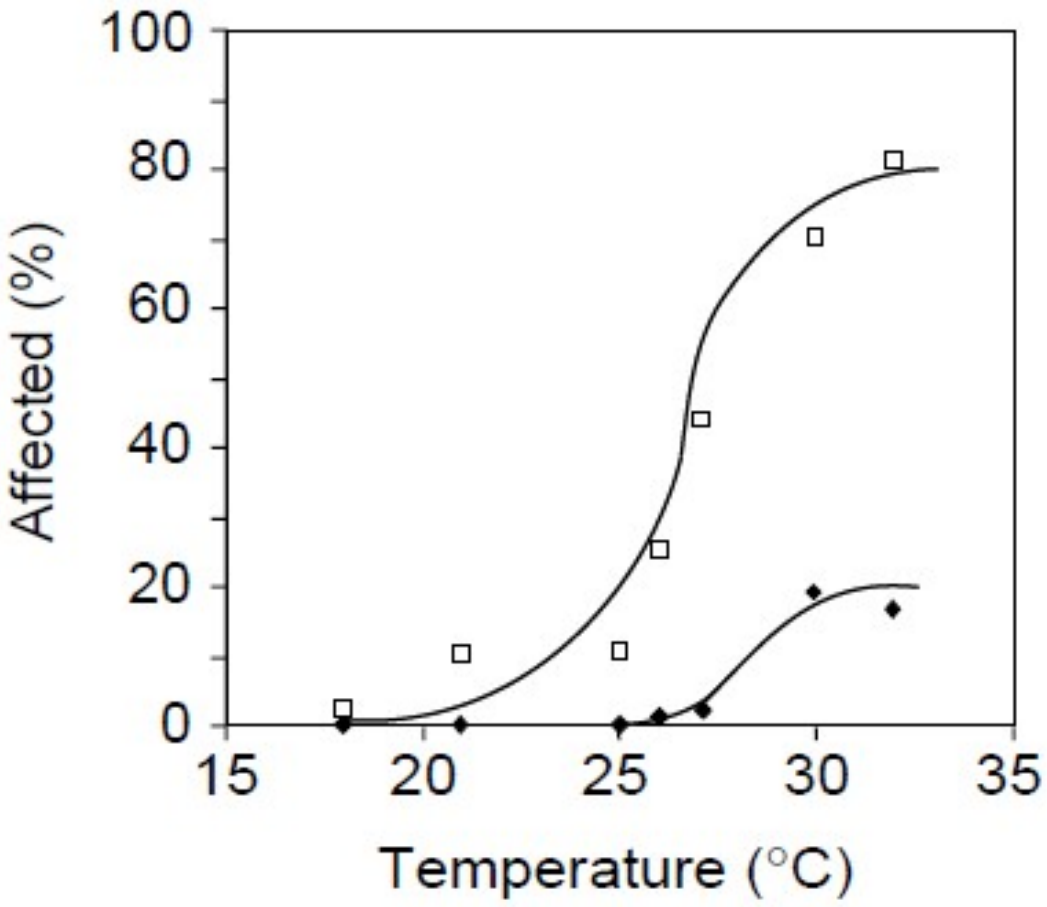


Figure 5. **Effect of *Hsp90* mutation on deformed eye trait in *Drosophila melangaster*.** Black diamonds indicate the wild type phenotype for deformed eye trait. Open squares indicate flies with a defective mutation in *Hsp90*. From Rutherford and Lindquist (1998).

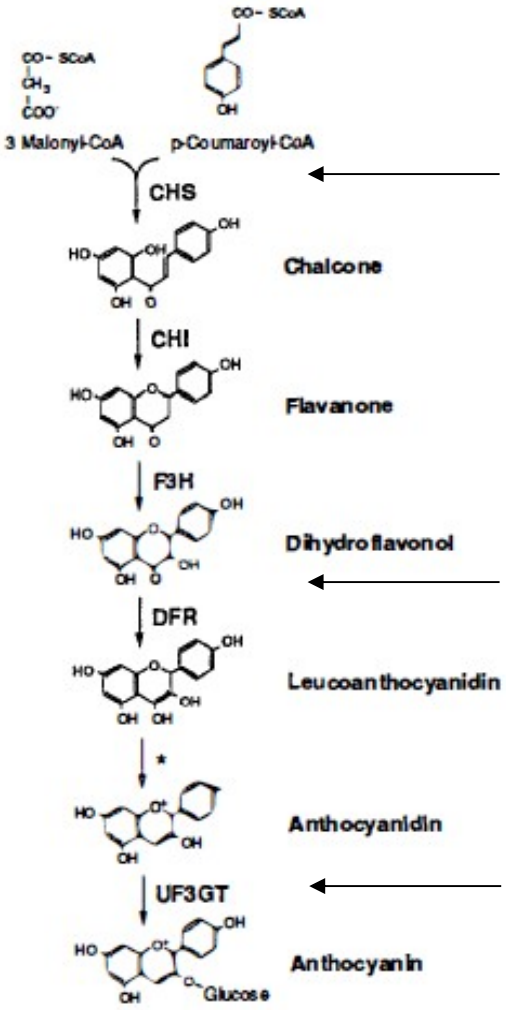


Figure 6. **Synthesis pathway of anthocyanin in *Zea mays*. The core linear synthesis pathway for glycosolated anthocyanins.** All biosynthetic enzymes of the pathway appear to be regulated regardless of tissue type, although the *r1* regulatory gene, which is partially regulated by the *vp1* gene (viviparous) (Light induced, black arrow), regulates the *Chalcone synthase* gene(CHS), *dihydroflavonol 4-reductase,*and *udp-glucose flavinoid 3-oxy-glucisyl transferase* in aleurone, scuteller node, and mesocotyl of young seedlings (horizontal black arrows) (Eggleston, W., pers. comm.). Enzymes listed but not regulated by *r1* include *CHI* (Chalcone isomerase) and *F3H* (Flavinoid 3`isomerase). From Dooner *et al.* (1992).

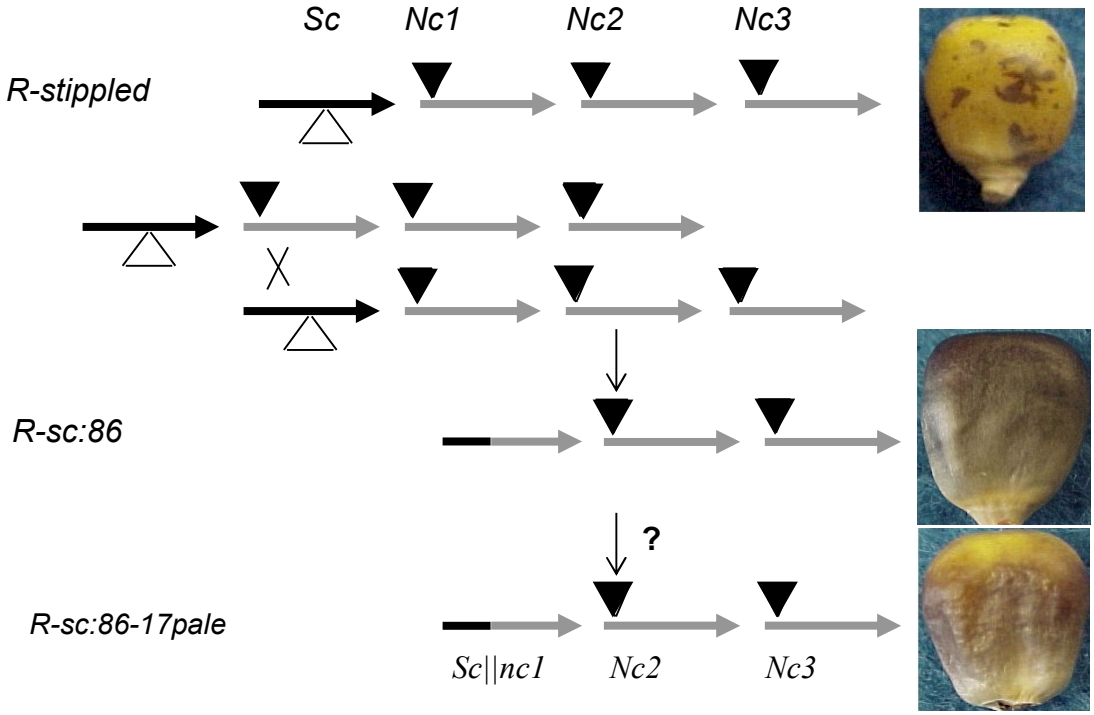


Figure 7. **Origin and molecular structure of *R-sc:86-17pale***. Horizontal arrows represent *r1* genes and direction of transcription. Centromere is to the left. Black arrows represent the *Sc* gene, shaded arrows represent the *Nc* genes and upright triangles represent the *I-R* transposable element inserted in exon 1 of *Sc*. *Doppia* elements in the 5' ends of *Nc1, Nc2,* and *Nc3* genes are inserted at bp +83 of the *Nc* genes shown as filled down-facing triangles (W. Eggleston, pers. comm.). The large X represents presumptive unequal crossover between *Sc* and *Nc1* to create the chimeric *Sc||nc1* gene in *R-sc:86*. Kernel phenotypes are shown to right (not to scale). Vertical arrows indicate mutational events in a parent line. From W. Eggleston (pers. comm.).

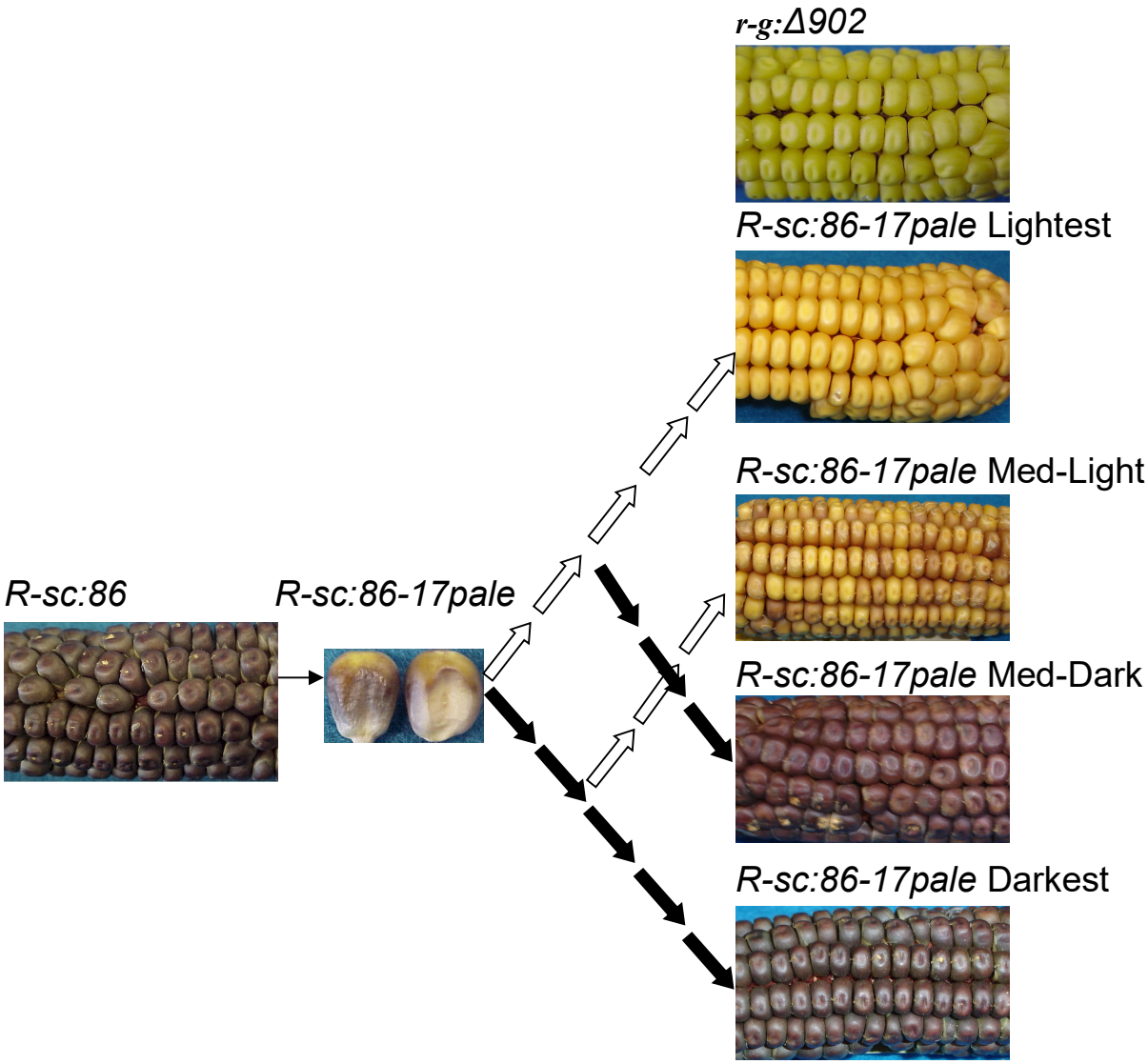


Figure 8. **Isolation of the sublines of *R-sc:86-17pale*.** The *r1* locus is deleted in *r-g:Δ902* and *R-sc:86* is the fully colored parent allele of *R-sc:86-17pale* (W. Eggleston, pers. comm.) (Eggleston *et al.* 1995) (Figure 7). Open arrows denote one generation of self-pollinated selection for lightest kernel color from an ear, while filled arrows denote one generation of selection for darkest kernels. From W. Eggleston (pers. comm.).

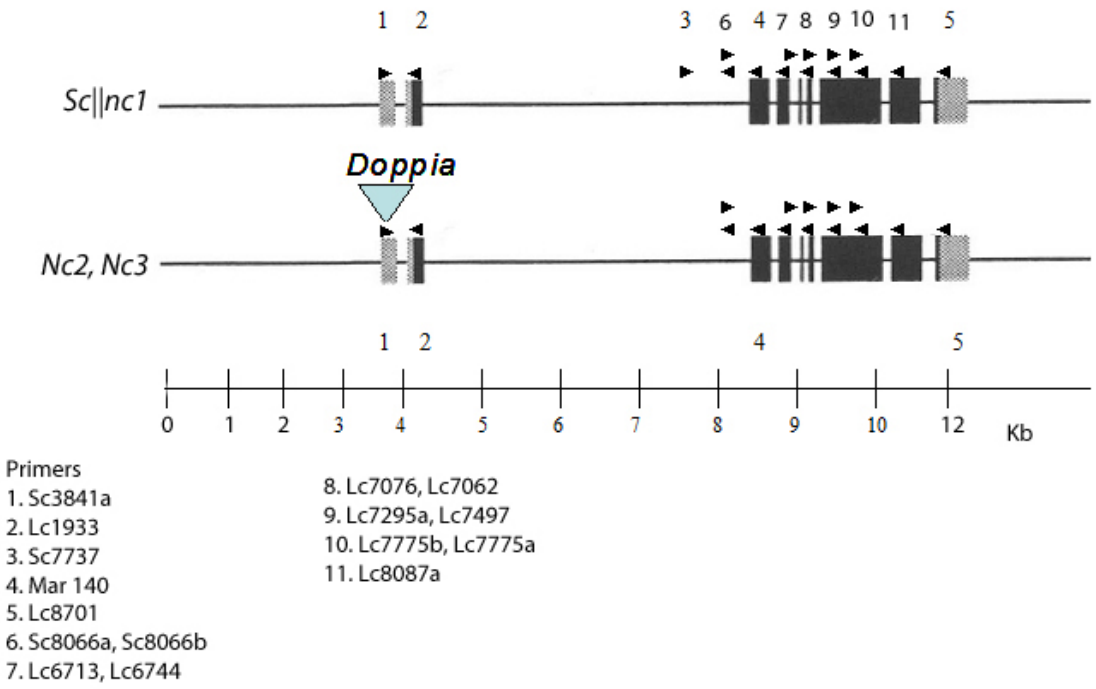


Figure 9. **Structure of *Sc||nc1, Nc2,* and *Nc3,* and oligonucleotides used for DNA sequencing.** The structure of the *Sc||nc1* gene and the structure of the *Nc2* & *Nc3* genes; The *Sc||nc1* gene is a chimera comprised of the *Sc* 5' end fused to the 3' end of the *Nc1* due to recombination in the 3' end of intron 2 (W. Eggleston, per. comm.). Horizontal arrows denote forward and reverse primers sequencing primers, while numbers above arrows correspond to key numbers.

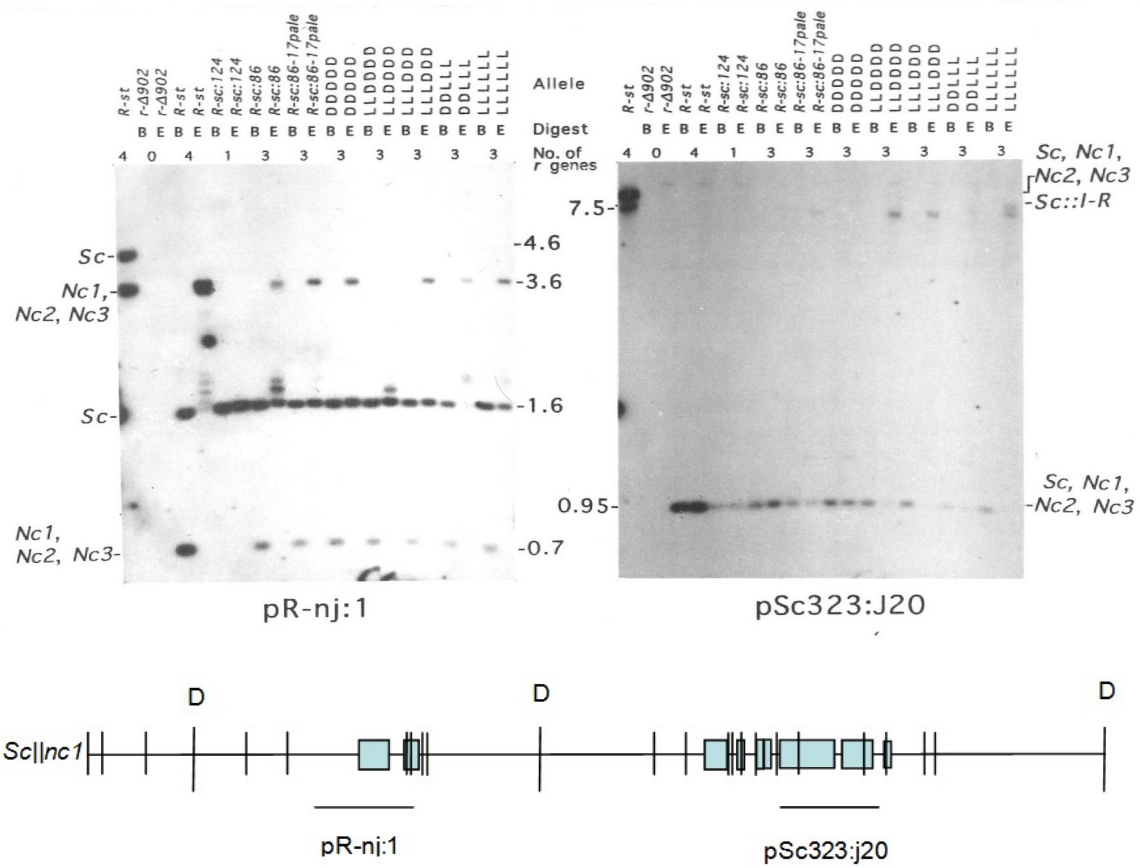


Figure 10. **Southern blot restriction map analysis of cytosine (CHG) methylation in *R-sc:86* and *R-sc:86-17pale* selection sublines**. DNA from seedlings homozygous for alleles noted were digested with *Hin*dIII (methylation insensitive) alone (left lane) or in combination with the isochzomers *Eco*RII (E) (sensitive to CHG methylation) or *Bst*NI (insensitive to CHG methylation). Left shows hybridization with pR-nj:1 and right, with pSc323:J20, with molecular weight (kb) in the middle (Eggleston *et al.* 1995). The pR-nj:1 blot probe can anneal to a fragment of up 4.3 kb in size once the DNA is digested with *Hin*dIII, and the pSc323:J20 blot probe can anneal to a fragment of up 7.5 kb in size once the DNA is digested with *Hin*dIII. Above each lane is the allele, enzyme and number of *r1* genes present. Probes are shown for a canonical *r1* gene in which transcription begins to the left. The thin line indicates introns, blue boxes indicate exons,

based on comparison of the *Sc* sequence (M. Alleman, J. Kermicle and W. Eggleston, pers. comm.) with *Lc* (S. Ludwig, L. Habera and S. Wessler, pers. comm.) (Ludwig *et al.* 1989; Ludwig *et al.* 1990). L's and D's as in Figure 8. *R-sc:124* contains only the single *r1* gene *Sc*||*nc3* (Eggleston *et al.* 1995). From W. Eggleston (pers. comm.). Vertical lines marked 'D' on the gene structure indicates *Hin*dIII recognition sites, while unmarked vertical lines on the gene structure indicate *Eco*RII/*Bst*NI recongition sites.

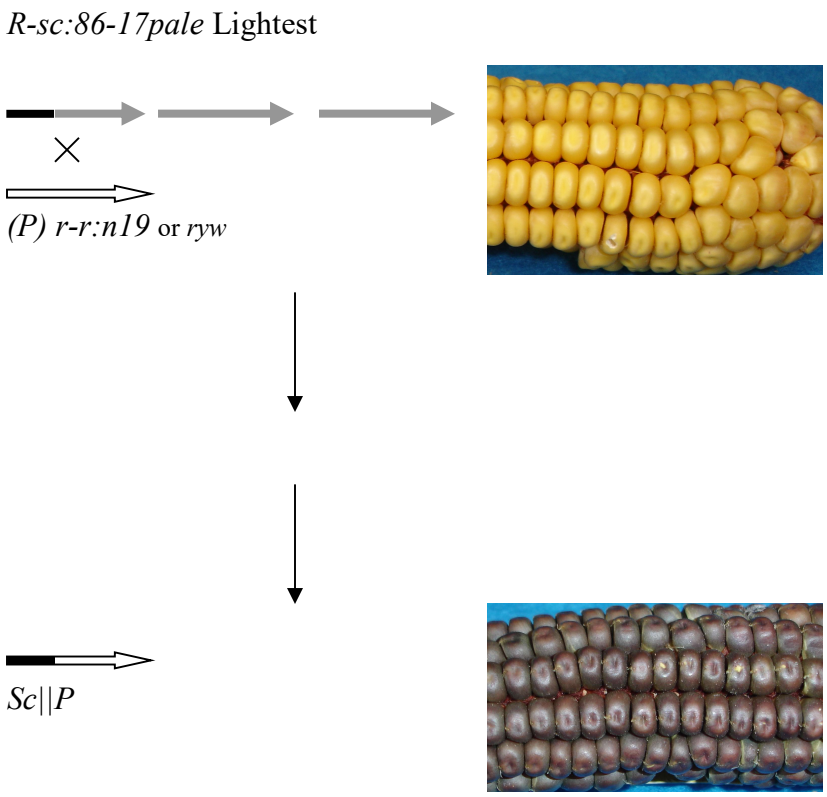


Figure 11. **Molecular basis for fully colored seed recovered from female plants heterozygous for *R-sc:86-17pale* Lightest and *r-r:n19* or *ryw; wx*+ crossed with *g R-r:8pale; wx-* males.** Arrows represent *r1* genes and direction of transcription. Black and grey arrows represent the *r1* genes and regions from R-*sc:86-17pale* Lightest, and white arrows represent *r1* genes from *r-r:n19* or *ryw*. From W. Eggleston (pers. comm.)

A) 

B) 

C) 

Figure 12. **Effect of *rmr6-* and *mop-* mutations on the *R-sc:86-17pale* Lightest kernel color.** A) An ear of *R-sc:86-17pale rmr6+/+; mop1* +/+ Lightest subline. B) Kernels from a *R-sc:86-17pale* Light; *rmr6-/-; mop1+/+*. C) An ear of *R-sc:86-17pale* Lightest; *rmr6+/+;mop1-/-*.

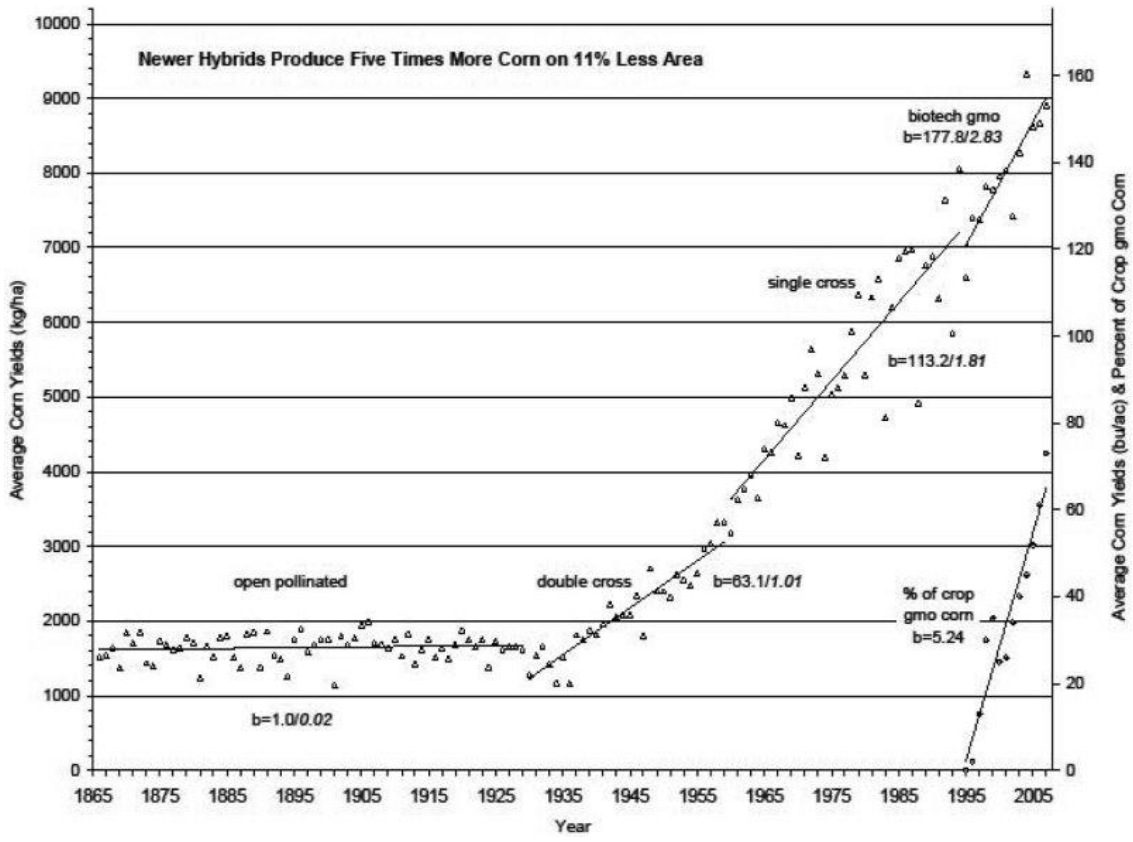


Figure 13. **Maize production in the US from 1865 to 2010.** Analysis of the historical annual yield per hectacre, with changes in yield as attributable to changes in breeding practices and the widespread adoption of hybrids and GMO varieties. From Phillips (2010).

Figure 14. **Kernels in the Agtron sample reading cup.** Darkest kernels from 2014 are shown in the clear Agtron sample cup, with more than 150 kernels in the cup to enable a valid light reflectance measurement. Kernels are at least two layers deep throughout the sample cup.

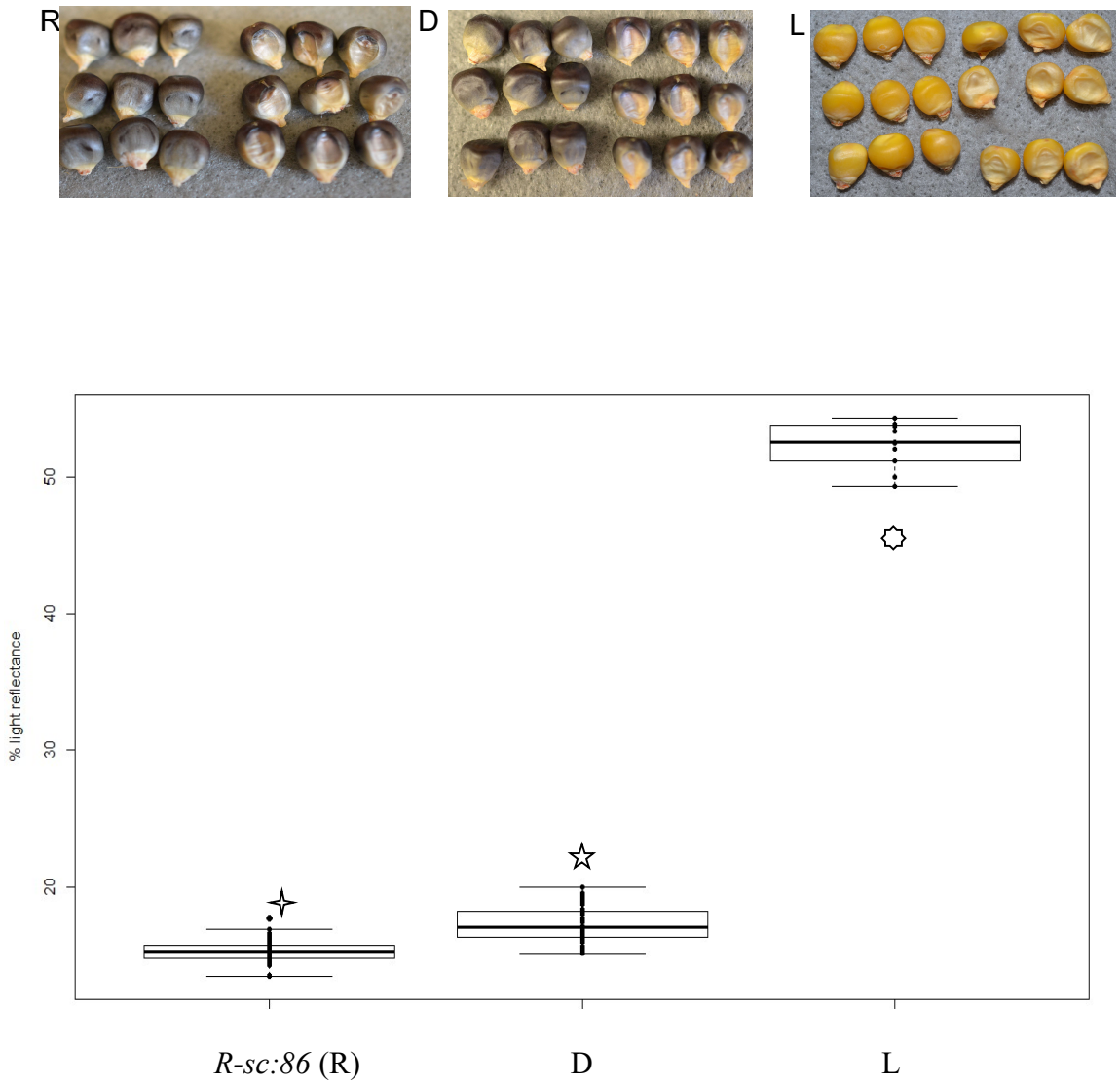


Figure 15. **Percent light reflectance of 2011 maize kernels by line.** R) Kernels from *R-sc:86.* D) Kernels from *R-sc:86-17pale* Darkest subline. L) Kernels from *R-sc:86-17pale* Lightest subline. Box plot of kernel light reflectance average measurements, with boxes indicating 1st and 3rd quartile of each sample set, samples shown as circles, and whiskers showing 2 standard deviations from the mean of each line and subline. Stars indicate statistically significant differences between lines tested (Table 3).

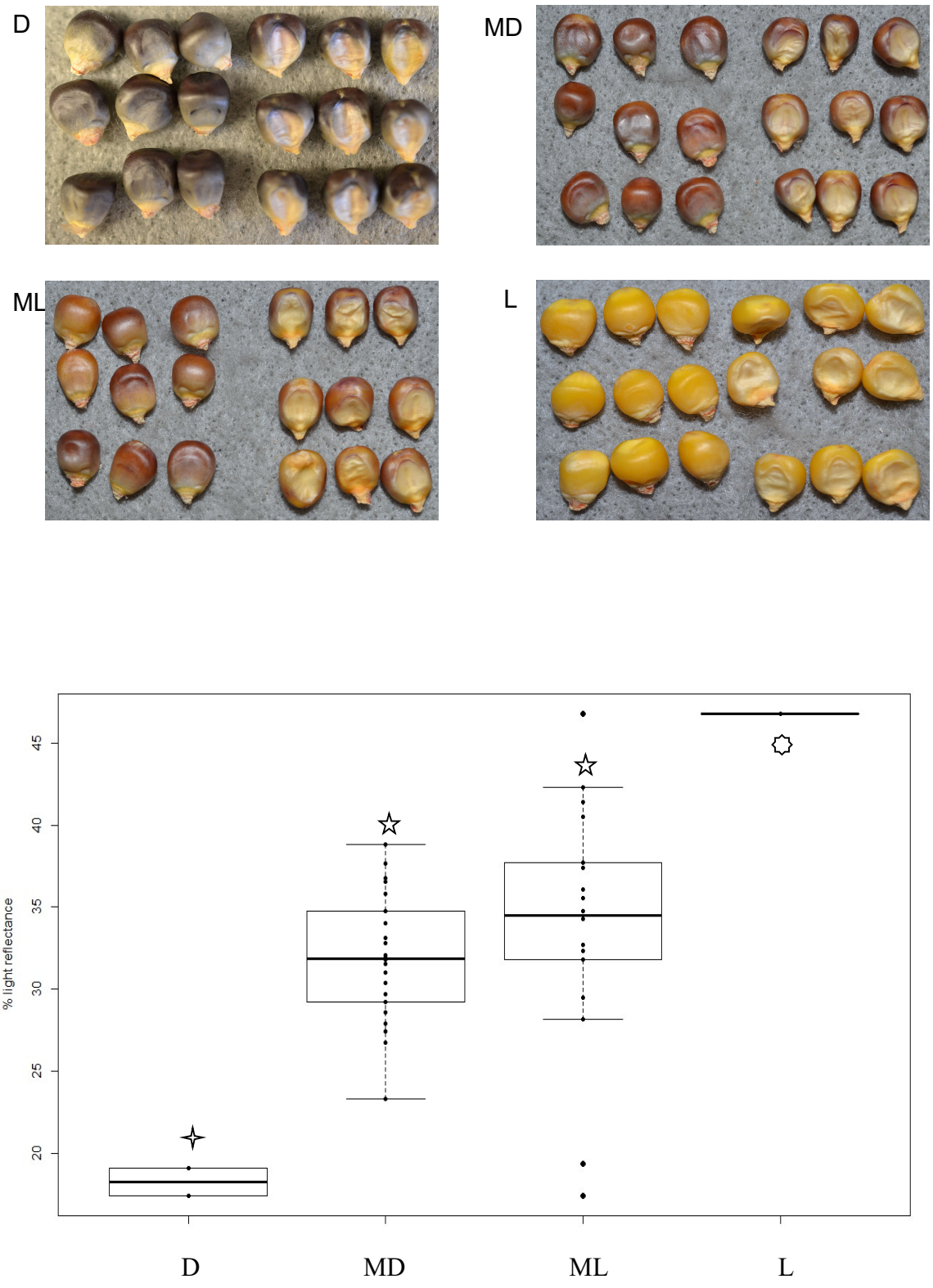


Figure 16. **Percent light reflectance of 2012 maize kernels by line.** D) Kernels from *R-sc:86-17pale* Darkest subline. MD) Kernels from *R-sc:86-17pale* Medium-Dark subline.

ML) Kernels from *R-sc:86-17pale* Medium-Light subline. L) Kernels from *R-sc:86-17pale* Lightest subline. Box plot of kernel light reflectance average measurements for each subline. Details as shown in Figure 15. Stars indicate statistically significant differences between lines tested (Table 4).

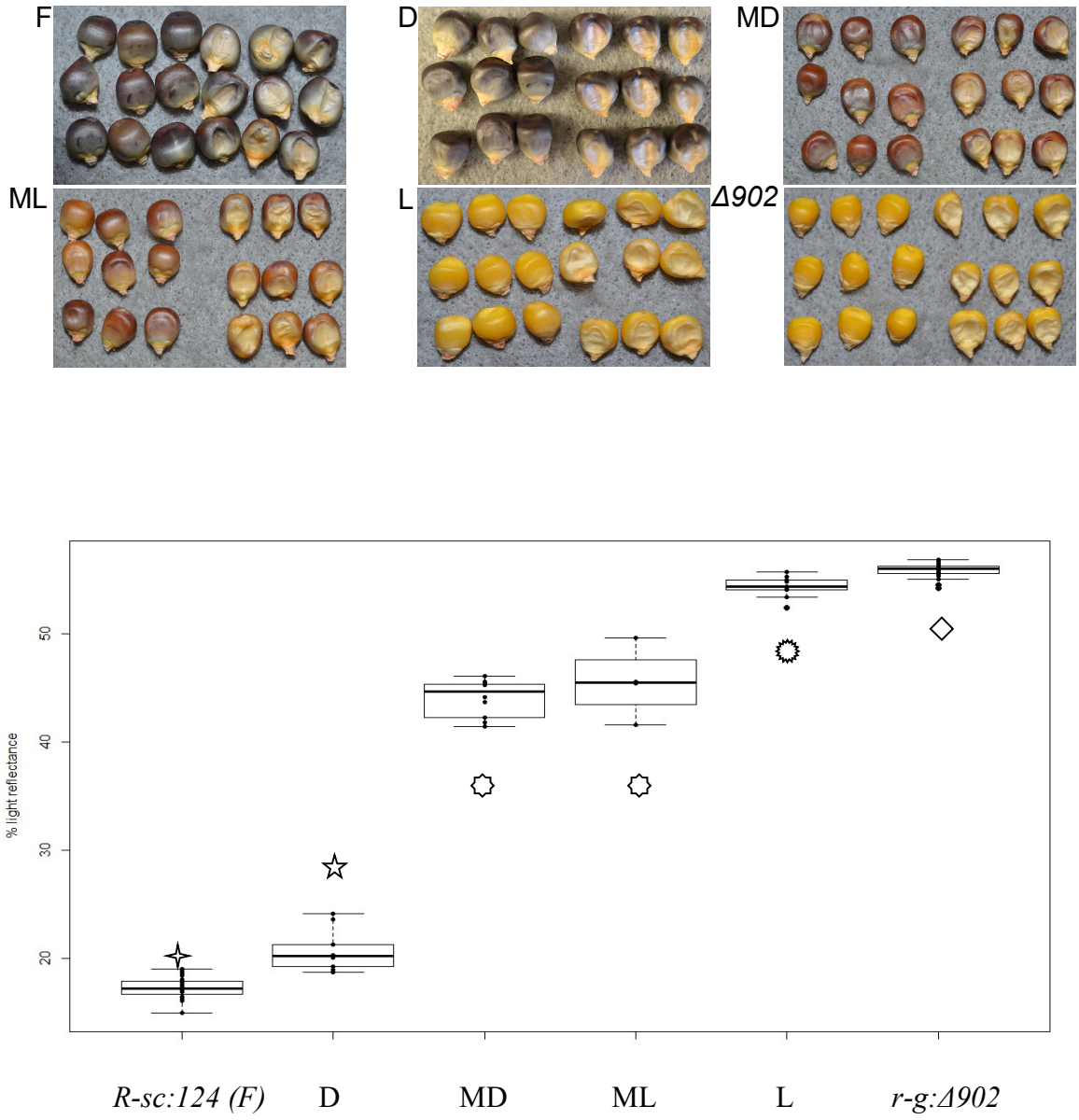


Figure 17. Percent light reflectance of 2014 maize kernels by line. F) Kernels from *R-sc:86.* D) Kernels from *R-sc:86-17pale* Darkest subline. MD) Kernels from *R-sc:86-17pale* Medium-Dark subline. ML) Kernels from *R-sc:86-17pale* Medium-Light subline. L) Kernels from *R-sc:86-17pale* Lightest subline. *Δ902*) Kernels from *r-g: Δ902* line. Box plot of kernel light reflectance average measurements for each line and subline. Details are as shown in Figure 15. Stars indicate statistically significant differences between lines tested (Table 5).

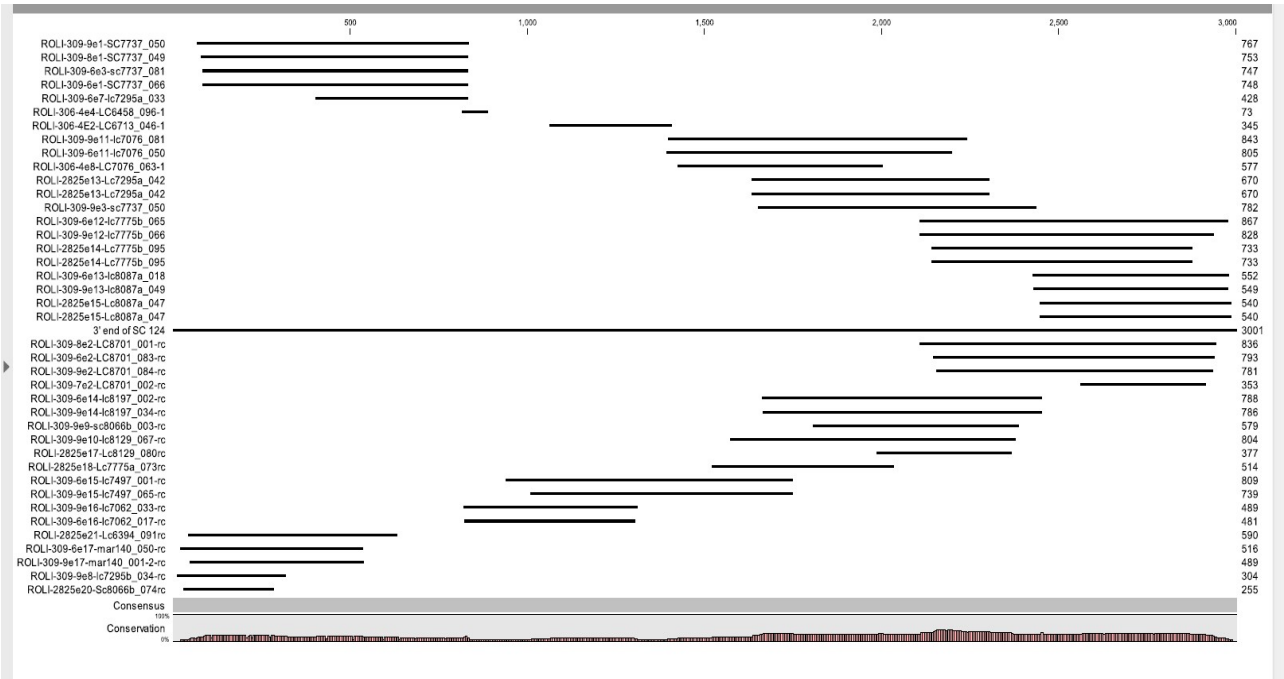


Figure 18. **Alignment of sequence reads from the Lightest subline to a reference sequence from *R-sc:124*.** The horizontal line in center represents the reference sequence from *R-sc:124* (W. Eggleston and M. Alleman, pers. comm.), and each of the shorter lines represents a sequence fragment aligned to create a consensus sequence. The sequence fragments above the reference line are fragments from the top backbone, and the sequence fragments from the bottom backbone strand. Each fragment is labeled to the left, with the sequencing primer used indicated.

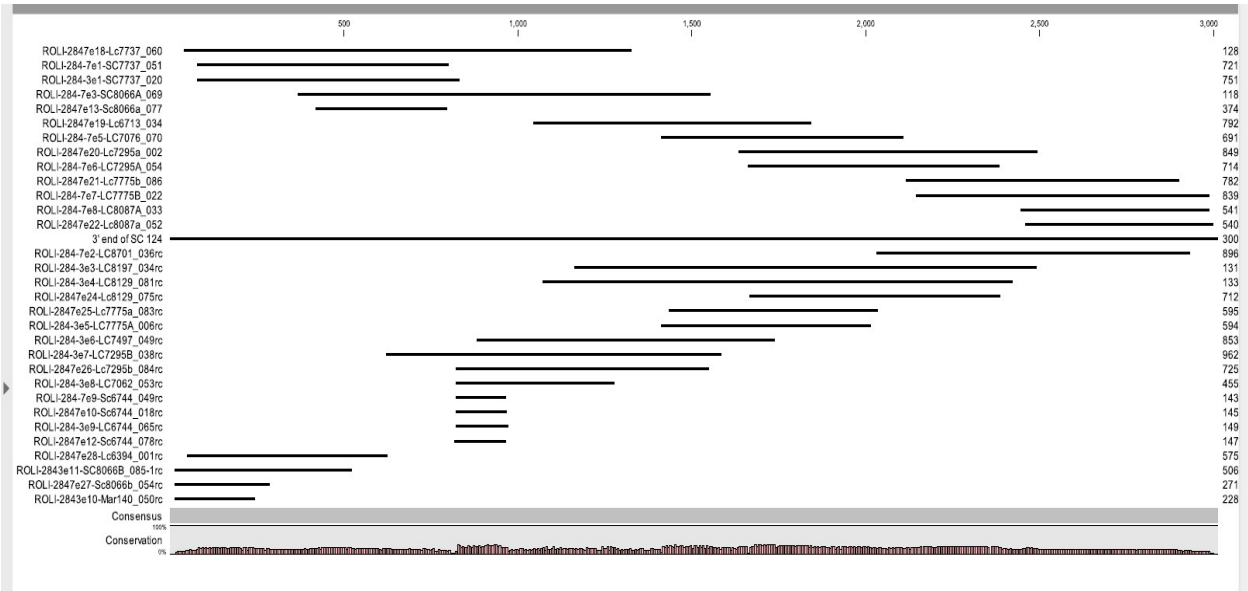


Figure 19. **Alignment of sequence reads from the Darkest subline to a reference sequence from *R-sc:124*.** The horizontal line in center represents the reference sequence from *R-sc:124* (W. Eggleston and M. Alleman, pers. comm.), and each of the shorter lines represents a sequence fragment aligned to create a consensus sequence. The sequence fragments above the reference line are fragments from the top backbone, and the sequence fragments from the bottom backbone strand. Each fragment is labeled to the left, with the sequencing primer used indicated.

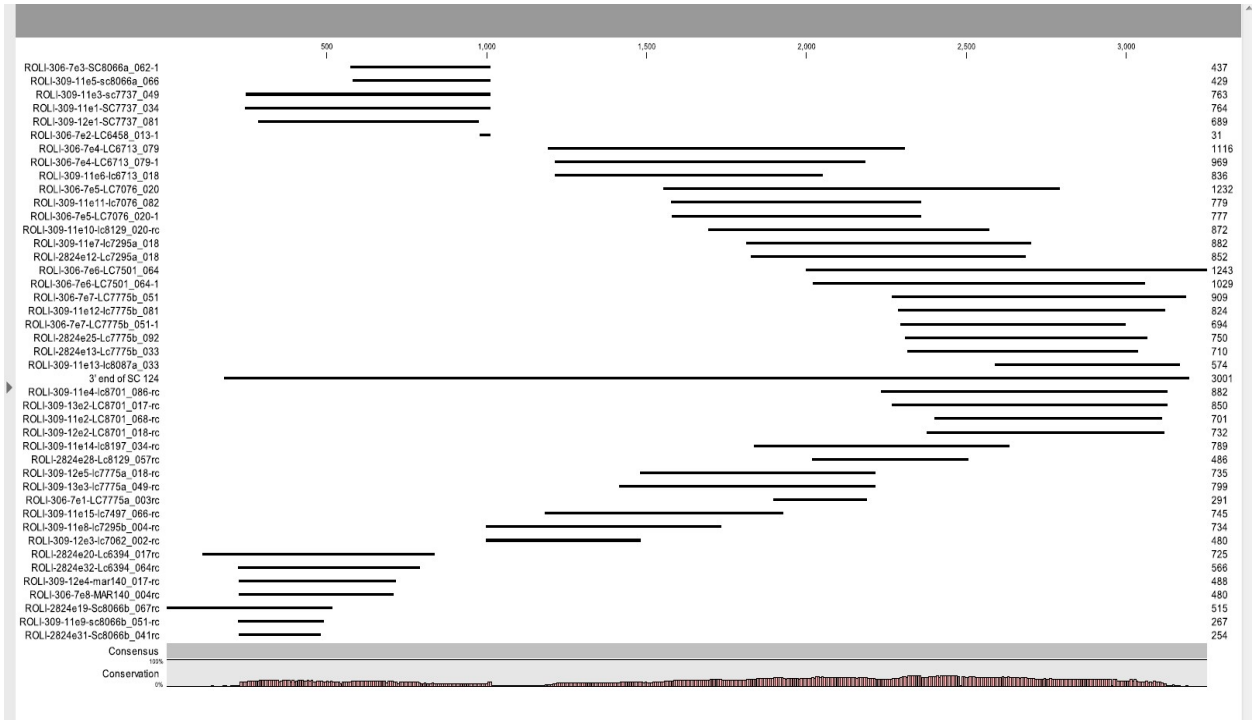
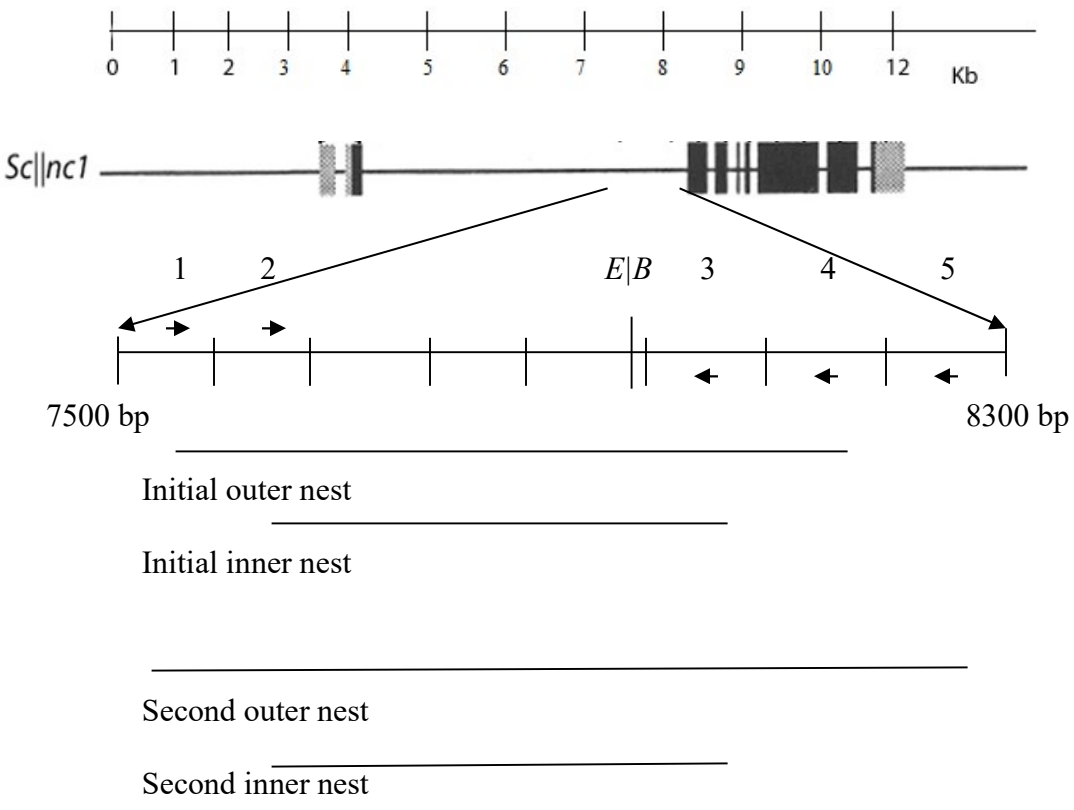


Figure 20. **Alignment of sequence reads from the progenitor line (*R-sc:86)* to a reference sequence from *R-sc:124*.** The horizontal line in center represents the reference sequence from *R-sc:124* (W. Eggleston and M. Alleman, pers. comm.), and each of the shorter lines represents a sequence fragment aligned to create a consensus sequence. The sequence fragments above the reference line are fragments from the top backbone, and the sequence fragments from the bottom backbone strand. Each fragment is labeled to the left, with the sequencing primer used indicated.



Primers and selected restriction sites
1. Bs47527
2. Bs47666m
3. Bs2Sc8045r
4. Bs2Sc8144r
5. Bs2Lc6237r
E|*B*. *Eco*RII/*Bst*N1 restriction site

Figure 21. **Structure of the *Sc*||*nc1* gene, with a blown up section showing the region amplified from genomic sodium bisulfite treated DNA**. Site 3 is a recognition sequence for the *Eco*RII/*Bst*N1 isoschizomers. The boundary between Intron 2 and Exon 3 is at 8273 bp, ~15 bp downstream of the Bs2Lc6237r primer. Horizontal bars show the amplicons from the first and second outer and inner nests, while forward and reverse arrows indicate primer annealing locations.

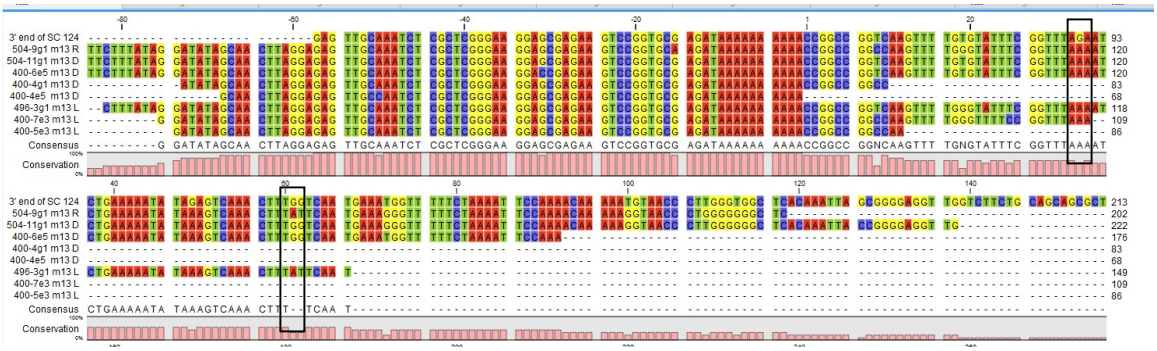


Figure 22. **Alignment of the sequence reads of the sodium bisulfite converted PCR amplicons for *R-sc:86*, *R-sc:86-17pale* Lightest, and *R-sc:86-17pale* Darkest to a reference sequence from *R-sc:124*.** The 3` end of SC 124 is the reference sequence used in Chapter 3 from *R-sc:124* (W. Eggleston pers. comm.), the second strand is the *R-sc:86* amplicon, and strands three through nine are the Darkest (D) and Lightest (L) amplicons. Boxes show a site with no methylation and a site with partial methylation.

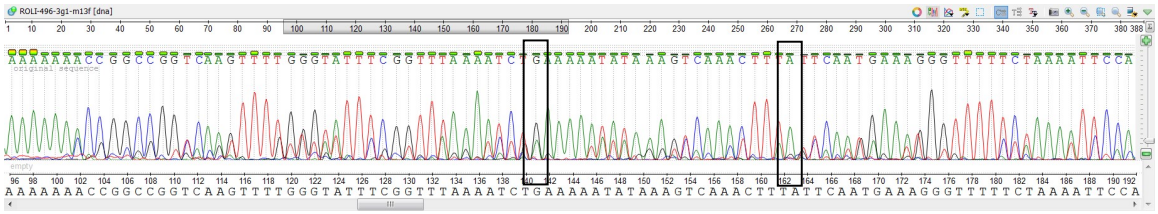


Figure 23**. Electropherograph of a sequenced amplicon of the *R-sc:86-17pale* Lightest subline, after treatment of genomic DNA with sodium bisulfite.** The first box shows a CHH site with partial methylation yielding both guanine and adenine peaks at the site. The second box shows a site that is fully unmethylated, showing only an adenine peak.

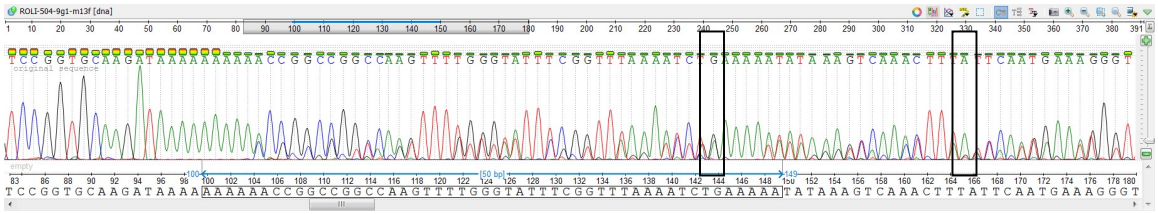


Figure 24**. Electropherograph of a sequenced amplicon of *R-sc:86* line, after treatment of genomic DNA with sodium bisulfite.** The first box shows a CHH site with partial methylation yielding both guanine and adenine peaks at the site. The second box shows a site that is fully unmethylated, showing only an adenine peak.

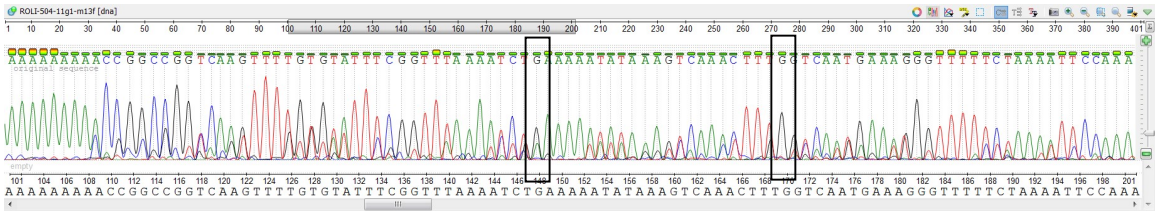


Figure 25. **Electropherograph of a sequenced amplicon of *R-sc:86-17pale* Darkest subline, after treatment of genomic DNA with sodium bisulfite.** The first box shows a CHH site with partial methylation yielding both guanine and adenine peaks at the site. The second box shows a site that is fully methylated, showing only a guanine peak.

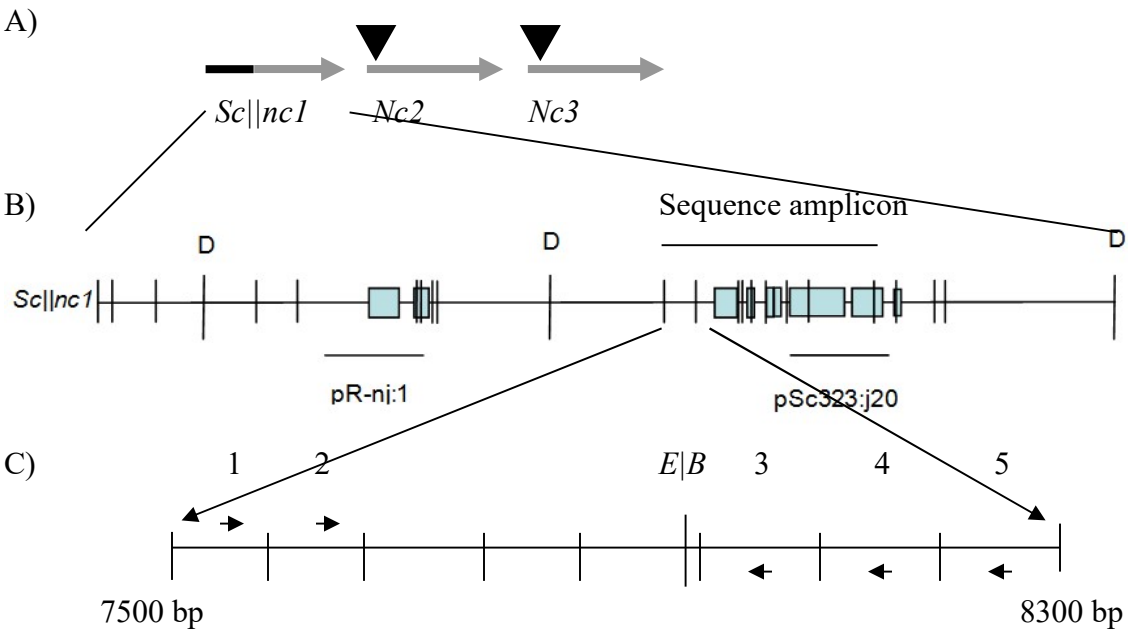


Figure 26. **Summary of the structure of *Sc||nc1*, the region sequenced, and the region analyzed for cytosine methylation patterns.** A) Shows the structure of the *r1* alleles (*Sc||nc1*, *Nc2,* and *Nc3*) present in *R-sc:86* and the *R-sc:86-17pale* selection sublines. B) An enlarged map of the *Sc||nc1* gene, with the Southern blot probe locations shown below the gene. *Hin*dIII restriction sites are shown by lines labeled 'D', with shorter vertical lines on the gene indicating recognition sites for the isochzomers *Eco*RII (methylation sensitive) and *Bst*NI (methylation insensitive). The horizontal bar labeled 'Sequence amplicon' is the region sequenced, and spans from 7737 bp to 10680 bp. C) The enlarged region of the *Sc||nc1* gene shows the region targeted for amplification using sodium bisulfite-specific primers. The region sequenced after amplification using these primers was from 7666 bp through 8045 bp, and includes a *Eco*RII|*Bst*N1 recognition site.

VITAE

Robert C. Lindsay was born in 1986 in Rock Hill, South Carolina. He graduated from Midlothian High school in Midlothian, Virginia in 2004. He attended Hood College and graduated with a Bachelor of Arts with a double major in Business Management and Biochemistry in 2008. Robert went on to the Horticultural Master's program at Virginia Polytechnic Institute and State University, where he was inducted into the Phi Sigma Biological Sciences Honor Society, and graduated with a Masters of Horticultural Science in 2010. Robert's Master's Thesis was entitled "Developing High-Throughput Techniques for the Generation of *F. Vesca* Mutants with an *Ac/Ds* insertional Element." Robert returned to Richmond, and adjuncted for the 2011 spring semester at Virginia Commonwealth University (VCU), and then entered the Integrated Life Sciences Ph.D. program at VCU in the fall of 2011. While at VCU, Robert has taught several different classes, and has been elected as President, Vice-President, and Member-at-Large in the Integrated Life Sciences Student Organization.